2018

Joint Analysis for Multiple Traits

Zhenchuan Wang
Michigan Technological University, zwang10@mtu.edu

JOINT ANALYSIS FOR MULTIPLE TRAITS

By

Zhenchuan Wang

A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

In Mathematical Sciences

MICHIGAN TECHNOLOGICAL UNIVERSITY

2018

This dissertation has been approved in partial fulfillment of the requirements for the Degree of DOCTOR OF PHILOSOPHY in Mathematical Sciences.

Department of Mathematical Sciences

Dissertation Co-Advisor:	*Shuanglin Zhang*
Dissertation Co-Advisor:	*Qiuying Sha*
Committee Member:	*Kui Zhang*
Committee Member:	*Laura E. Brown*
Department Chair:	*Mark S. Gockenbach*

Table of Contents

List of tables

List of figures

Preface

This dissertation is submitted for the degree of Doctor of Philosophy at Michigan Technological University. The research described herein was conducted under the supervision of Prof. Shuanglin Zhang and Prof. Qiuying Sha in the Department of Mathematical Sciences, Michigan Technological University, between September 2013 and March 2018.

This work is to the best of my knowledge original, except where references are made to previous work. Part of this work contains previously published material. The title of Chapter 1 is *Joint analysis of multiple traits in rare variant association studies* and it was published in the Annals of Human Genetics (Wang, Z., Wang, X., Sha, Q., and Zhang, S., 2016, 80(3):162-171). The overall study was designed by Shuanglin Zhang. Zhenchuan Wang and Shuanglin Zhang conducted the statistical analyses. Zhenchuan Wang, Shuanglin Zhang, and Qiuying Sha drafted the manuscript. Chapter 2 entitled *Joint Analysis of Multiple Traits Using "Optimal" Maximum Heritability Test was published* in Plos One (Wang, Z., Sha, Q., and Zhang, S. 2016, 11(3): e0150975). Zhenchuan Wang, Shuanglin Zhang, and Qiuying Sha developed the methodology and wrote the original draft. Zhenchuan Wang and Shuanglin Zhang performed formal analysis. Shuanglin Zhang and Qiuying Sha provided comments and review for editing. The title of Chapter 3 is *Testing an optimally weighted combination of common and/or rare variants with multiple traits* and it is under review in PLOS ONE (Wang, Z., Sha, Q., Zhang K., and Zhang, S. 2018). Shuanglin Zhang designed the overall study. Zhenchuan Wang, Shuanglin Zhang, and Kui Zhang conducted statistical analyses, and Zhenchuan Wang, Qiuying sha and Shaunglin Zhang wrote the manuscript.

Acknowledgements

At Michigan Technological University, I have obtained valuable experience in massive testing and high-dimensional data and analysis by effective collaborations with scientists in statistical genetics and biotechnology. I would like to express my deep appreciations to many professionals who helped me broaden my knowledge.

First of all, I express my sincere gratitude to the committee. I owe a great debt of gratitude to Dr. Shuanglin Zhang, who serves as my advisor and deserves my special appreciation for his dedicated advising and long-term support. Forever, I am motivated by his exponential insights and efforts in statistical genetics. I extend my appreciation to Dr. Sha for her important advice and discussions on my research topics, to Dr. Kui Zhang and Dr. Laura Brown who carefully examined my dissertation and suggested substantial improvements. It is my honor to have these outstanding professors on my committee.

Many influential graduate students and professors deserve special thanks for their collaborations and instructions. I cannot name them all but must thank Dr. Xuexia Wang, Huanhuan Zhu, Dr. Xiao Zhang, Dr. Jianping Dong, and Dr. Mark S. Gockenbach.

Abstract

This dissertation includes three papers with each distributed in one chapter.

In chapter 1, we proposed an Adaptive Weighting Reverse Regression (AWRR) method to test association between multiple traits and rare variants in a genomic region. AWRR is robust to the directions of effects of causal variants and is also robust to the directions of association of traits. Using extensive simulation studies, we compared the performance of AWRR with canonical correlation analysis (CCA), Single-TOW, and the Weighted Sum Reverse Regression (WSRR). Our results showed that, in all of the simulation scenarios, AWRR is consistently more powerful than CCA. In most scenarios, AWRR is more powerful than Single-TOW and WSRR.

In chapter 2, we proposed an "optimal" maximum heritability test (MHT-O) to test the association between multiple traits and a single variant. MHT-O includes a procedure of deleting traits that have weak or no association with the variant. Using extensive simulation studies, we compared the performance of MHT-O with MHT, Trait-based Association Test uses Extended Simes procedure (TATES), SUM_SCORE and MANOVA. Our results showed that, in all of the simulation scenarios, MHT-O is either the most powerful test or comparable to the most powerful test among the five tests we compared.

In chapter 3, we developed a statistical method by testing an optimally weighted combination of variants with multiple traits (TOWmuT) to test the association between multiple traits and a weighted combination of variants (rare and/or common) in a genomic region. TOWmuT is robust to the directions of effects of causal variants and is applicable to different types of traits. Using extensive simulation studies, we compared the performance of TOWmuT with the following five existing methods: gene association with multiple traits (GAMuT), multiple sequence kernel association test (MSKAT), adaptive weighting reverse regression (AWRR), single-TOW, and MANOVA. Our results showed that, in all of the simulation scenarios, TOWmuT has correct type I error rates and is consistently more powerful than the other five tests. We also illustrated the usefulness of TOWmuT by analyzing a whole-genome genotyping data from a lung function study.

1 Chapter 1

Joint analysis of multiple traits in rare variant association studies

Abstract: The joint analysis of multiple traits has recently become popular since it can increase statistical power to detect genetic variants and there is increasing evidence showing that pleiotropy is a widespread phenomenon in complex diseases. Currently, most of existing methods for the joint analysis of multiple traits are to test association between one common variant and multiple traits. However, the variant-by-variant methods for common variant association studies may not be optimal for rare variant association studies due to the allelic heterogeneity as well as the extreme rarity of individual variants. Current statistical methods for rare variant association studies are for one single trait only. In this paper, we propose an Adaptive Weighting Reverse Regression (AWRR) method to test association between multiple traits and rare variants in a genomic region. AWRR is robust to the directions of effects of causal variants and is also robust to the directions of association of traits. Using extensive simulation studies, we compare the performance of AWRR with canonical correlation analysis (CCA), Single-TOW, and the Weighted Sum Reverse Regression (WSRR). Our results show that, in all of the simulation scenarios, AWRR is consistently more powerful than CCA. In most scenarios, AWRR is more powerful than Single-TOW and WSRR.

Introduction

There is increasing evidence showing that pleiotropy, the effect of one variant on multiple traits, is a widespread phenomenon in complex diseases [Sivakumaran et al., 2011]. Furthermore, in genetic association studies of complex diseases, multiple related traits are usually measured. For example, hypertension is evaluated using systolic and diastolic blood pressures, the Metabolic Syndrome is based on observing three of five criteria [Sattar et al., 2008], and neuropsychiatric disorders depend on a range of overlapping clinical characteristics[O'Reilly et al., 2012]. Although most published genome-wide association studies (GWAS) analyze each of the related traits separately, the joint analysis of multiple traits not only can increase statistical power to detect genetic variants [Solovieff et al., 2013; Stephens, 2013; Yang & Wang, 2012; Zhou & Stephens, 2014], but also can be crucial to understand the genetic architecture of the disease of interest [Aschard et al., 2014]. Thus, the joint analysis of multiple traits has recently become popular. Several statistical methods for the joint analysis of multiple traits have been developed. These methods can be roughly divided into three groups: regression methods, combining test statistics from univariate analysis, and dimension reduction methods. Regression methods include mixed effects models [Korte et al., 2012; Zhou & Stephens, 2014] and reverse regression models [O'Reilly et al., 2012; Yan et al., 2013]. By modeling the covariance structure of correlated traits and dependence structure between individuals, mixed effects models not only can incorporate multiple correlated traits, but also can be robust to population stratification. Reverse regression models consider genotypes as the response variable and all the traits as independent variables, therefore, reverse regression models do

not need to know the complex distributions of the traits and can be applied to a large number of mixed types of traits. For combining the test statistics from univariate analysis, one first obtains univariate test statistics by performing association tests for each trait individually and then combines the univariate test statistics by linear combinations [O'Brien, 1984; van der Sluis et al., 2013; Yang et al., 2010]. The dimension reduction methods include canonical correlation analysis (CCA) [Tang & Ferreira, 2012], principal components of traits (PCT) [Aschard et al., 2014], and principal components of heritability (PCH) [Klei et al., 2008; Lange et al., 2004; Ott & Rabinowitz, 1999]. CCA is to find a linear combination of traits and a linear combination of genotypes at multiple variants such that the correlation between the two linear combinations reaches its maximum. PCT is the principal component analysis to the traits. The PCT methods are usually based on the first PC or first few PCs of the traits [Feng et al., 2007; Klei et al., 2008]. Aschard et al. (2014) showed that contrary to the widespread practice, tests based on only the first few PCs often have low power, whereas combining signals across all PCs can have greater power. PCH is to find a linear combination of multiple traits such that this linear combination has the maximum heritability.

Almost all of the aforementioned methods are to test association between one common variant and multiple traits. However, the variant-by-variant methods for common variant association studies may not be optimal for rare variant association studies due to the allelic heterogeneity as well as the extreme rarity of individual variants [Li & Leal, 2008]. Recent studies show that complex diseases are caused by both common and rare variants [Bodmer & Bonilla, 2008; Kang et al., 2010; Pritchard, 2001; Pritchard & Cox, 2002; Stratton & Rahman, 2008; Teer & Mullikin, 2010; Walsh & King, 2007]. Next-generation sequencing technology allows sequencing of the whole genome of large groups of individuals, and thus makes rare variant association studies feasible [Andres et al., 2007; Metzker, 2010]. Recently, statistical methods for rare variant association studies with a single trait have been developed by summarizing genotype information from multiple variants. These methods include burden tests [Li & Leal, 2008; Madsen & Browning, 2009; Morgenthaler & Thilly, 2007; Price et al., 2010; Zawistowski et al., 2010], quadratic tests [Neale et al., 2011; Sha et al., 2012; Wu et al., 2011], and combined tests [Derkach et al., 2013; Lee et al., 2012; Sha & Zhang, 2014]. Burden tests collapse rare variants in a genomic region into a single burden variable and then regress the trait on the burden variable to test for the cumulative effects of rare variants in the region. These tests implicitly assume that all rare variants are causal and that the directions of the effects are all the same. Quadratic tests include tests with statistics of quadratic form of score vector, as well as adaptive weighting methods. These tests are robust to the directions of the effects of causal variants and are less affected by neutral variants than burden tests. Burden tests can only outperform quadratic tests when most of rare variants are causal and the directions of the effects of causal variants are all the same. Combined tests combine information from burden tests, quadratic tests, and possibly other tests aiming to have advantages of multiple tests.

In this article, we propose an adaptive weighting reverse regression (AWRR) method to test association between multiple traits and rare variants in a genomic region. In

AWRR, we first propose adaptive weights to collapse genotypes. Then, we use the score test to test association based on the reverse regression, in which collapsed genotypes is treated as the response variable and multiple traits are treated as independent variables. Using extensive simulation studies, we compare the performance of AWRR with CCA, Single-TOW, and the Weighted Sum Reverse Regression (WSRR). In the Single-TOW, we first calculate the TOW statistic [Sha et al., 2012] to test the association between each trait and variants in a genomic region and then the statistic of Single-TOW is the largest of TOW statistics. In the WSRR, we first calculate the weighted sum [Madsen & Browning, 2009] of genotypes at variants in a genomic region and then the statistic of WSRR is the score test statistic under reverse regression model, in which the weighted sum of genotypes is the response variable and traits are predictor variables. Our results show that, in all of the simulation scenarios, AWRR is consistently more powerful than CCA. In most scenarios, AWRR is more powerful than Single-TOW and WSRR.

Methods

We consider a sample with n unrelated individuals. Each individual has K (potentially correlated) traits and has been genotyped at M variants in a genomic region. Let y_{ik} denote the k^{th} trait value of the i^{th} individual. Let x_{im} denote the genotype score of the m^{th} variant of the i^{th} individual, where x_{im} is the number of minor alleles of the i^{th} individual carried at the m^{th} variant. We denote $Y_i = (y_{i1},\ldots,y_{iK})^T$ as the K traits for the i^{th} individual. We propose an adaptive weighting reverse regression (AWRR) method to test the null hypothesis H_0: none of the K traits are associated with the M variants in the genomic region. For constructing the test statistic of AWRR, we first collapse the M dimensional genotype $(x_{i1},\ldots,x_{iM})$ into a one dimensional number $x_i = \sum_{m=1}^{M} w_m x_{im}$, where w_m is the adaptive weight for the m^{th} variant. The adaptive weight w_m should satisfy properties that w_m should be large if the m^{th} variant has strong association with the K traits and w_m should have different signs for risk and protective variants. Then, the statistic of AWRR is the score test statistic under the reverse regression model $x_i = \beta_0 + \sum_{k=1}^{K} \beta_k y_{ik} + \varepsilon_i$. In details, the AWRR method has the following steps.

1. We define a weight W_m for the m^{th} variant such that W_m will be large if the m^{th} variant has strong association with the K traits and W_m will be also large if the m^{th} variant is a rare variant. For these purposes, we propose $W_m = \frac{1}{\sqrt{p_m(1-p_m)}} T_m$, where p_m is the minor allele frequency of the m^{th} variant and T_m is the score statistic to test the null hypothesis $H_0: \beta_1 = \cdots = \beta_K = 0$ under the reverse regression model $\log\frac{p_{im}}{1-p_{im}} = \beta_0 + \beta_1 y_{i1} + + \beta_K y_{iK}$, where we assume a dominant model $p_{im} = \Pr(x_{im}=1) = \Pr(x_{im}=2)$. In fact, for rare variants, x_{im} essentially is 0 or 1. The score statistic is given by $T_m = U_m^T V_m^{-1} U_m$, where $U_m = \sum_{i=1}^{n} Y_i (x_{im} - \bar{x}_m)$ and $V_m = \frac{1}{n}\sum_{i=1}^{n}(x_{im}-\bar{x}_m)^2 \sum_{i=1}^{n}(Y_i - \bar{Y})(Y_i - \bar{Y})^T$, where $\bar{x}_m = \frac{1}{n}\sum_{i=1}^{n} x_{im}$, and $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$. Under the null hypothesis, T_m follows a χ^2 distribution with degrees of freedom K. However, W_m does not consider the direction of the effects of causal variants.
2. In this step, we will define a direction of W_m. We first select a trait (the selected trait denoted as the $\tilde{k}$th trait). We use $sign(\rho(y_{\tilde{k}}, x_m))$ to denote the direction of the associations of the m^{th} variant, where $y_{\tilde{k}} = (y_{1\tilde{k}}, \ldots, y_{n\tilde{k}})$ and

$x_m = (x_{1m},\ldots,x_{nm})^T$. If the $\tilde{k}$th trait has no association with the M variants, the directions of the association will be random. In order to try to avoid random directions, we propose to choose the trait that has the strongest association with the M variants. Let T^k_{TOW} denote the statistic of TOW [Sha et al., 2012] to test association between the k^{th} trait and the M variants. T^k_{TOW} is defined as $T^k_{TOW} = \sum_{i=1}^{n}(y_{ik} - \bar{y}_k)(x_i^o - \bar{x}^o)$, where $x_i^o = \sum_{m=1}^{M} w_m^o x_{im}$ and $w_m^O = \sum_{i=1}^{n}(y_{ik} - \bar{y}_k)(x_{im} - \bar{x}_m) / \sum_{i=1}^{n}(x_{im} - \bar{x}_m)^2$. We choose the $\tilde{k}$th trait such that $T^{\tilde{k}}_{TOW} = \max_{1\le k\le K} T^k_{TOW}$.

3. The final weight for the m^{th} variant is given by $w_m = \text{sign}\left(\rho\left(y_{\tilde{k}}, x_m\right)\right)W_m$. Let $x_i = \sum_{m=1}^{M} w_m x_{im}$. Then, we consider the reverse regression model $x_i = \beta_0 + \sum_{k=1}^{K}\beta_k y_{ik} + \varepsilon_i$. We apply a score test to test the null hypothesis $H_0: \beta_1 = \cdots = \beta_K = 0$. The score statistic is given by $T_{AWRR} = u^T V^{-1} u$, where $u = \sum_{i=1}^{n}(x_i - \bar{x})Y_i$, $V = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})(Y_i - \bar{Y})^T$, $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, and $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$.
4. In this step, we evaluate the p-value of T_{AWRR}. Since w_m depends on the trait values and the genotype scores, T_{AWRR} does not follow a χ^2 distribution with degrees of freedom K. We use a permutation procedure to evaluate the p-value of T_{AWRR}. In each permutation, we randomly shuffle $Y_1,\ldots,Y_n$ and keep the genotypes of each individual unchanged. We repeat step 1 to step 3 based on each permutated data. Let T^0_{AWRR} denote the test statistic of T_{AWRR} based on the original data and T^{per}_{AWRR} denote the test statistic based on the permuted data. Then the p-value of the test T_{AWRR} is the proportion of the number of permutations with $T^{per}_{AWRR} \ge T^0_{AWRR}$.

Comparisons of Tests

We compare the performance of the proposed test AWRR with those of the canonical correlation analysis (CCA) [Tang & Ferreira, 2012], the Single-TOW method [Sha et al., 2012], and the Weighted Sum Reverse Regression (WSRR) method [Madsen & Browning, 2009].

CCA method: although the asymptotical distribution of the CCA statistic works well for common variants, it is very conservative for rare variants. Thus, we propose to use a permutation procedure to evaluate the p-value of the CCA statistic.

Single-TOW method: let T_{TOW}^{k} denote the statistic of TOW [Sha et al., 2012] to test association between the k^{th} trait and the M variants. The statistic of Single-TOW is given by $T_{Single-TOW} = \max_{1 \le k \le K} T_{TOW}^{k}$. The p-value of $T_{Single-TOW}$ is evaluated by a permutation procedure.

WSRR method: let $X_i = \sum_{m=1}^{M} w_m x_{im}$, where $w_m = 1/\sqrt{p_m(1-p_m)}$ and p_m is the minor allele frequency of the m^{th} variant. We consider the reverse regression model $X_i = \beta_0 + \sum_{k=1}^{K} \beta_k y_{ik} + \varepsilon_i$. The statistic of WSRR, T_{WSRR}, is the score test statistic to test the null hypothesis $H_0 : \beta_1 = \mathsf{L} = \beta_K = 0$. Under the null hypothesis, T_{WSRR} follows a χ^2 distribution with degrees of freedom K.

Simulation Study

Our simulations follow that of Sha et al. (2012). In details, the empirical Mini-Exome genotype data provided by the genetic analysis workshop 17 (GAW17) is used for simulation studies. This dataset contains genotypes of 697 unrelated individuals on 3205 genes. We will conduct two sets of simulations. In the first set of simulations, we choose four genes: ELAVL4, MSH4, PDE4B, and ADAMTS4 with 10, 20, 30, and 40 variants, respectively. We merge the four genes to form a super gene (Sgene1) with 100 variants [Sha et al., 2012]. In the second set of simulations, we choose ten genes: ELAVL4, FAM73A, PSMB4, FSHR, GMCL1, HNMT, GALNT13, NEUROD1, MYEOV2, and TWF2 with 10 variants in each of them. We merge the ten genes to form a super gene (Sgene2) with 100 variants. In our simulation studies, we generate genotypes based on the genotypes of 697 individuals in the Sgene1 and Sgene2. To generate a qualitative disease affection status, we use a liability threshold model based on a quantitative trait. For a qualitative trait, an individual is defined to be affected if the individual's corresponding quantitative trait is at least one standard deviation larger than the phenotypic mean. This yields a prevalence of 16% for the simulated disease in the general population. In the following, we describe how to generate a quantitative trait.

To evaluate the type I error, we generate K traits of an individual independent of the genotypes by using

$$Y = (y_1,\ldots,y_K)^T = \sqrt{\rho}\,\mathrm{B}u + \sqrt{1-\rho}\,\varepsilon,$$

where $u = (u_1,\ldots,u_{n_u})^T \sim MVN(0,I)$ is a vector of n_u independent standard normal latent variables, $\varepsilon = (\varepsilon_1,\ldots,\varepsilon_K)^T \sim MVN(0,I)$ is a vector of errors, B is a $K \times n_u$ loading matrix, the values of n_u and B are based on two variance models: (1) $n_u = 1$, $\mathrm{B} = (1,\ldots,1)^T$ and (2) $n_u = 2$, $\mathrm{B} = \begin{bmatrix} e_{[K/2]} & 0 \\ 0 & e_{K-[K/2]} \end{bmatrix}$. Thus, $Y \sim MVN(0,\Sigma)$, where $\Sigma = \rho\mathrm{B}\mathrm{B}^T + (1-\rho)I$.

To evaluate power, we consider that all causal variants are rare (MAF<0.01). We randomly choose n_c rare variants as causal variants, where n_c is determined by the percentage of causal variants among rare variants. Denote n_r and n_p as the number of risk rare variants and protective rare variants, respectively, where $n_r + n_p = n_c$. Let x_{qi}^r and x_{ji}^p denote the genotypic scores of the q^{th} risk rare variant and the j^{th} protective rare variant for the i^{th} individual, respectively. Suppose that causal variants have impact on the L traits among the K traits and, among the L traits, there are L_p traits positively correlated with risk variants and there are L_n traits negatively correlated with risk variants. Let h denote the heritability of all the n_c rare causal variants on each of the L traits. Generate n_c random numbers $r_1,\ldots,r_{n_c}$ from an uniform distribution between 0 and 1. The heritability

of the i^{th} causal variant is given by $h_i = hr_i \Big/ \sum_{j=1}^{n_c} r_j$. Under this assumption, we simulated K traits by

$$y_{ik} = \begin{cases} \sum_{q=1}^{n_r} \beta_{kq}^r x_{qi}^r - \sum_{j=1}^{n_p} \beta_{kj}^p x_{ji}^p + \varepsilon_{ik}, & 1 \le k \le L_p \\ -\left(\sum_{q=1}^{n_r} \beta_{kq}^r x_{qi}^r - \sum_{j=1}^{n_p} \beta_{kj}^p x_{ji}^p \right) + \varepsilon_{ik}, & L_p + 1 \le k \le L, \\ \varepsilon_{ik}, & L < k \le K \end{cases}$$

where $\varepsilon_i = \left(\varepsilon_{i1}, \ldots, \varepsilon_{iK}\right)^T$ can be generated in the same way as generating traits of evaluating type I error, β_{kq}^r and β_{kj}^p are constants and their values depend on the heritability.

Results

For type I error evaluation, we only consider the first set of simulations, but consider different sample sizes, different significance levels, different variance models and different types of traits. In each simulation scenario, the p-values of AWRR, Single-TOW and CCA are estimated by 10,000 permutations (the p-values of WSRR are estimated by a χ^2 distribution) and the type I error rates of all of the four tests are evaluated using 10,000 replicated samples. For 10,000 replicated samples, the 95% confidence intervals (CIs) for the type I error rates at the nominal levels 0.05 and 0.01 are (0.046, 0.054) and (0.008, 0.012), respectively. The estimated type I error rates of the four tests are summarized in Tables 1.2, 2.2, A.1.1 and A.1.2. From these tables, we can see that only two estimated type I error rates of CCA are not within the CIs and these two type I error rates (one is 0.0126 for nominal level 0.01 in Table 1.1, and the other one is 0.05505 for nominal level 0.05 in Table S1) are very close to the upper bounds of the corresponding CIs, which indicates that the four tests are all valid.

For power comparisons, we consider 10 traits and we assume that all causal variants are rare. For each type of traits and each variance model, we consider different values of heritability, different percentages of protective variants, and different percentages of causal variants. In each of the simulation scenarios, the p-values of AWRR, Single-TOW and CCA are estimated using 1,000 permutations (the p-values of WSRR are estimated by a χ^2 distribution) and the power of all of the four tests is evaluated using 500 replicated samples at a significance level of 0.05.

We first consider the first set of simulations for quantitative traits under variance model 1. Figure 1.1 provides the power comparisons of the four tests (WSRR, AWRR, CCA and Single-TOW) for the power as a function of heritability. This figure shows that WSRR is the least powerful one and AWRR is the most powerful one. It is little complicated to compare the power of Single-TOW with the power of CCA. When genotypes impact on only one trait, Single-TOW is more powerful than CCA; otherwise, CCA is more powerful than Single-TOW. Since Single-TOW only depends on the trait that has the strongest association with genotypes, it is more favorable for Single-TOW when genotypes impact on less traits. Power comparisons of the four tests for the power as a function of percentage of protective variants are given by Figure 1.2. This figure shows that, with the increasing of the percentage of protective variants, the power of WSRR decreases while the power of the other three methods does not change. Other patterns of the power comparisons are similar to those shown in Figure 1.1. The power comparisons of the four tests for the power as a function of the percentage of causal variants are given by Figure 1.3. As shown in this figure, with the increasing of the percentage of causal variants, the power of WSRR increases while the power of the other three methods does not change. WSRR is the least powerful one when the percentage of causal variants is small (≤ 0.15), while WSRR is the most powerful test when the percentage of causal variants is large (≥ 0.3). The patterns of the power comparisons of CCA, AWRR and Single-TOW are similar to those shown in Figure 1.1.

Under the first set of simulations, we also compare the powers of the four methods for quantitative traits under variance model 2 and for qualitative traits under variance models 1 and 2. These results are given in Figures B.1.1-B.1.9. For each type of traits, the patterns of the power comparisons are similar under variance models 1 and 2. For qualitative traits, CCA is consistently less powerful than Single-TOW and AWRR because CCA is designed for quantitative traits. For qualitative traits, the powers of AWRR, Single-TOW, and CCA decrease with the increase of the percentage of protective variants, although decrease not as fast as that of WSRR. As pointed out by Wu et al. (2011) and Sha et al. (2012), the decrease in the powers of AWRR, Single-TOW, and CCA in the presence of both risk and protective variants is due to the fact that protective variants lower MAFs in cases and thus make observing rare variants in the cases more difficult. The larger decrease in power of WSRR is additionally driven by the sensitivity to the direction of the effect due to aggregation of genotypes.

Under the second set of simulations, we compare the powers of the four methods for quantitative traits under variance model 1. Results are given in Figures S10-S12. Comparing Figures B.1.10-B.1.12 with Figures 1.1-1.3, the patterns of the power comparisons under the second set of simulations are very similar to that under the first set of simulations. Under the second set of simulations, we also compare the powers of the four methods for quantitative traits under variance model 2 and for qualitative traits under variance models 1 and 2 (results are not shown). Results also show that the patterns of the power comparisons under the second set of simulations are very similar to that under the first set of simulations.

In summary, for all simulation scenarios, AWRR is consistently more powerful than CCA and the power of WSRR increases with the increasing of the percentage of causal variants or with the decreasing of the percentage of protective variants. For quantitative traits, the powers of AWRR, CCA and Single-TOW are robust to the percentage of protective variants and to the percentage of causal variants, while for qualitative traits, the powers of AWRR, CCA and Single-TOW decrease with the increasing of the percentage of protective variants and are relatively robust to the percentage of causal variants.

Discussion

In this article, we proposed the AWRR method to perform joint analysis of multiple traits in rare variant association studies based on the following reasons: (1) the development of next-generation sequencing technology has made directly testing all rare variants feasible and (2) there is increasing evidence showing that pleiotropy is a widespread phenomenon in complex diseases and multiple related traits are usually measured in genetic association studies of complex diseases. We used extensive simulation studies to compare the performance of AWRR with CCA, WSRR and Single-TOW. Our results showed that AWRR has correct type I error rates, is robust to the directions of the association of causal variants for quantitative traits, and is robust to the percentage of causal variants. AWRR is consistently more powerful than CCA. AWRR is more powerful than Single-TOW and WSRR in most simulation scenarios.

Our simulation studies showed that the performance of each of AWRR, WSRR and Single-TOW depends strongly upon the number of traits impacted by genetic variants, the percentage of protective variants, and the percentage of causal variants. And no method demonstrates consistently good power. To increase the robustness of the test, we can combine AWRR, WSRR and Single-TOW aiming to have advantages of the three methods. Let p_{AWRR}, p_{WSRR} and $p_{Single-TOW}$ denote the p-values of AWRR, WSRR, and Single-TOW, respectively. The combined test statistic can be defined as $T_{combined} = \min\left\{p_{AWRR}, p_{WSRR}, p_{Single-TOW}\right\}$. However, the performance of the combined test needs further investigations.

In association studies based on unrelated individuals, it has been long recognized that population stratification can seriously confound association results [Knowler et al., 1988; Lander & Schork, 1994]. Several methods have been developed to control for population stratification for association studies based on unrelated individuals. These methods include GC approach [Devlin & Roeder, 1999; Devlin et al., 2001; Reich & Goldstein, 2001], PC approach [Bauchet et al., 2007; Chen et al., 2003; Price et al., 2006; Zhang et al., 2003; Zhu et al., 2002], and MLM approach [Kang et al., 2010; Zhang et al., 2010]. Like most association tests based on unrelated individuals, AWRR subjects to bias due to population stratification. To make AWRR robust to population stratification, we can use the PC approach. Let $P_i = \left(p_{i1}, \ldots, p_{iL}\right)^T$ denote the first L PCs of the genotypes at a set of genomic markers for the i^{th} individual. In step 3 of AWRR, we can use the residuals of the regression $x_i = \alpha + \beta^T P_i + \varepsilon_i$ to replace x_i and use the residuals of the regression $y_{ik} = \alpha_k + \beta_k^T P_i + \varepsilon_{ik}$ to replace y_{ik}. The performance of using the PC approach to control for population stratification in AWRR also needs further investigations.

The computation time required for running AWRR depends on the sample size, the number of variants in the genomic region, the number of traits, and the number of permutations. The running time of AWRR with 1000 permutations on the data set with 1000 individuals, 10 traits, and 100 variants in the genomic region on a laptop with 4 Intel

Cores @ 2.00GHz and 4 GB memory is no more than 0.5s. To perform genome-wide studies, we can first select genomic regions that show evidence of association based on a small number of permutations (e.g. 1,000), and then a large number of permutations are used to test the selected regions.

Tables and Figures

Table 1.1. The estimated type I error rates of four methods for quantitative traits under variance model 1. 10,000 replicates are used. This set of simulations is based on Sgene1.

	Sample size			
		500	1000	2000
$\alpha = 0.05$	CCA	0.0518	0.0519	0.04645
	Single-TOW	0.04995	0.05255	0.0506
	WSRR	0.0464	0.0506	0.0496
	AWRR	0.0519	0.0527	0.0531
$\alpha = 0.01$	CCA	0.012	0.00845	0.0126
	Single-TOW	0.0117	0.00965	0.012
	WSRR	0.0081	0.009	0.0097
	AWRR	0.01075	0.0097	0.01135

Table 1.2. The estimated type I error rates of four methods for qualitative traits under variance model 1. 10,000 replicates are used. This set of simulations is based on Sgene1.

	Sample size			
		500	1000	2000
$\alpha = 0.05$	CCA	0.052	0.0527	0.04985
	Single-TOW	0.0519	0.0534	0.05
	WSRR	0.0502	0.0493	0.0487
	AWRR	0.054	0.0505	0.05265
$\alpha = 0.01$	CCA	0.0101	0.01115	0.00955
	Single-TOW	0.0109	0.01165	0.01115
	WSRR	0.0106	0.0092	0.0106
	AWRR	0.00955	0.00965	0.012

Figure 1.1. Power comparisons of the four tests (WSRR, AWRR, CCA and Single-TOW) for the power as a function of heritability for quantitative traits under variance model 1. The sample size is 1000 and $\rho = 0.5$. The percentage of the causal variants is 0.1. All causal variants are risk variants. The total number of traits is 10. This set of simulations is based on Sgene1.

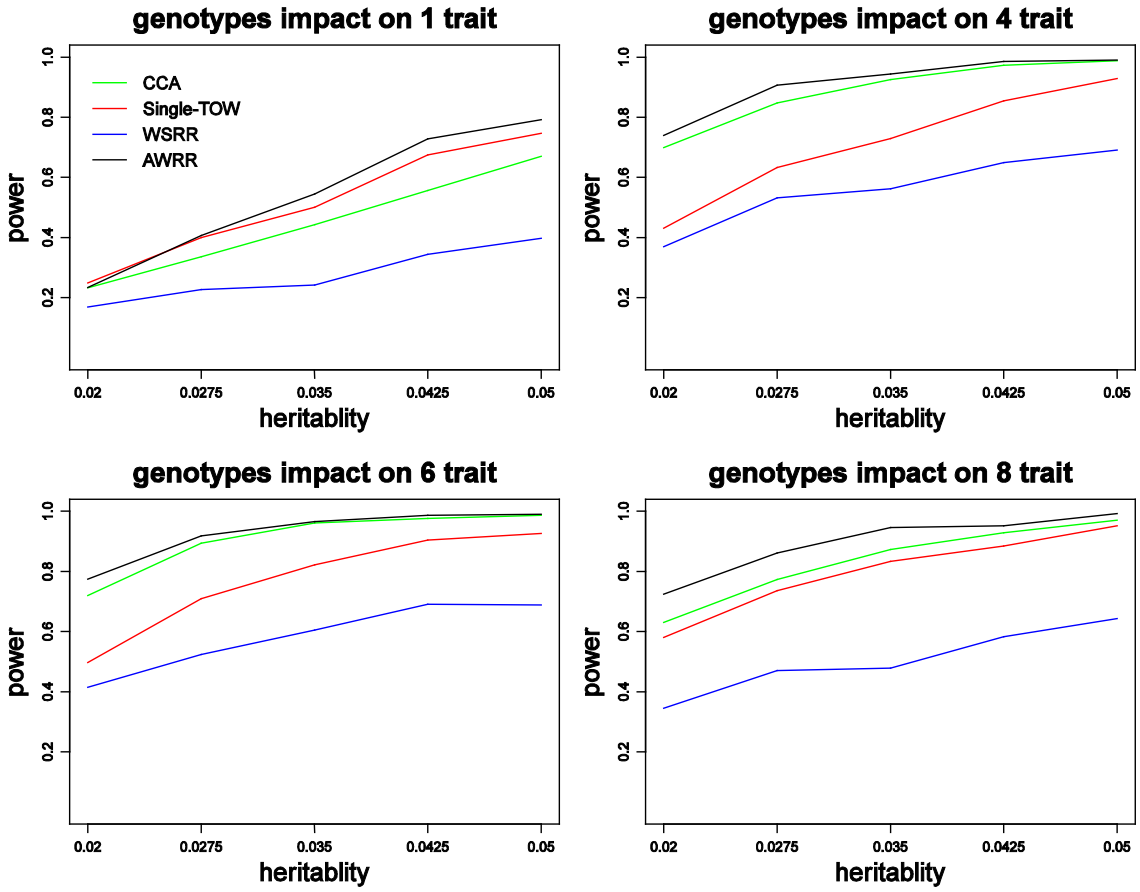


Figure 1.2. Power comparisons of the four tests (WSRR, AWRR, CCA and Single-TOW) for the power as a function of percentage of protective variants for quantitative traits under variance model 1. The sample size is 1000, the percentage of causal variants is 0.2, the total heritability is 0.03, and $\rho = 0.5$. The total number of traits is 10. This set of simulations is based on Sgene1.

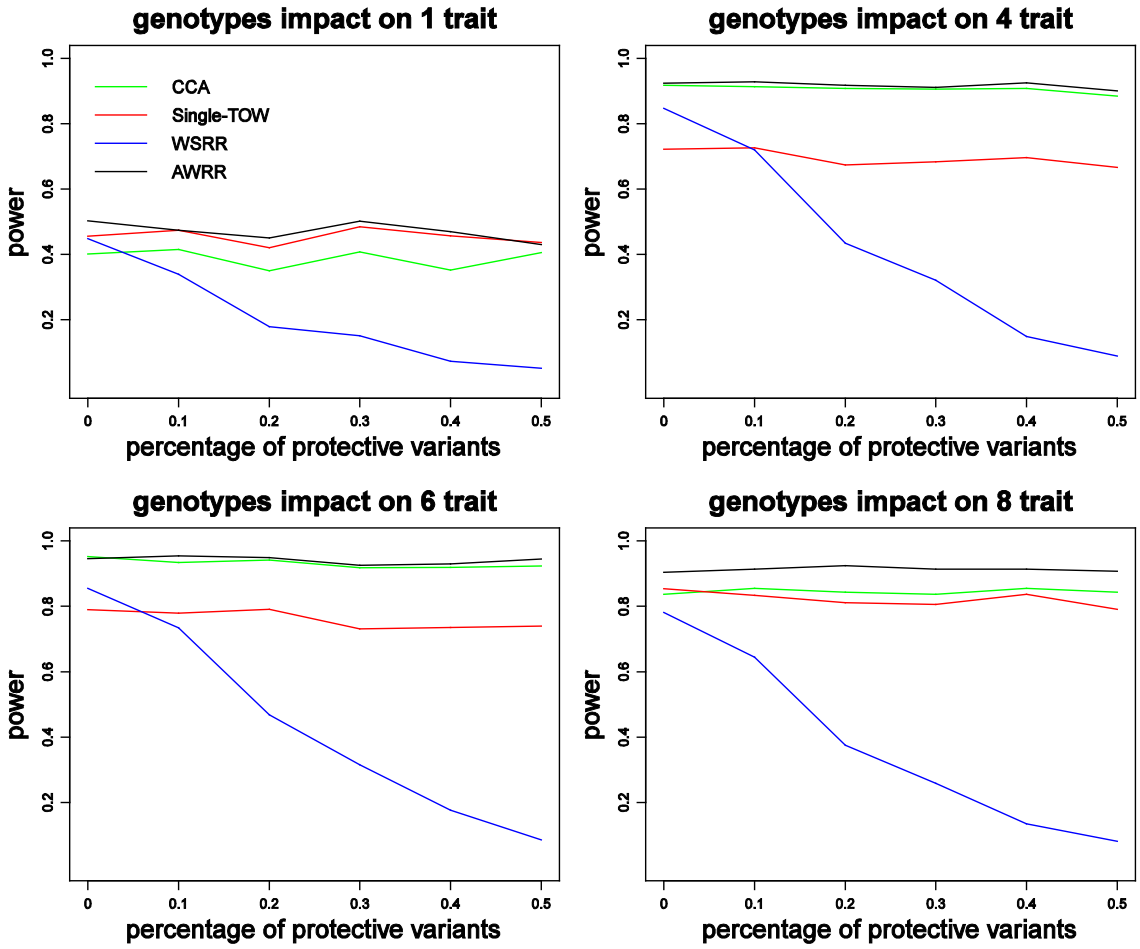
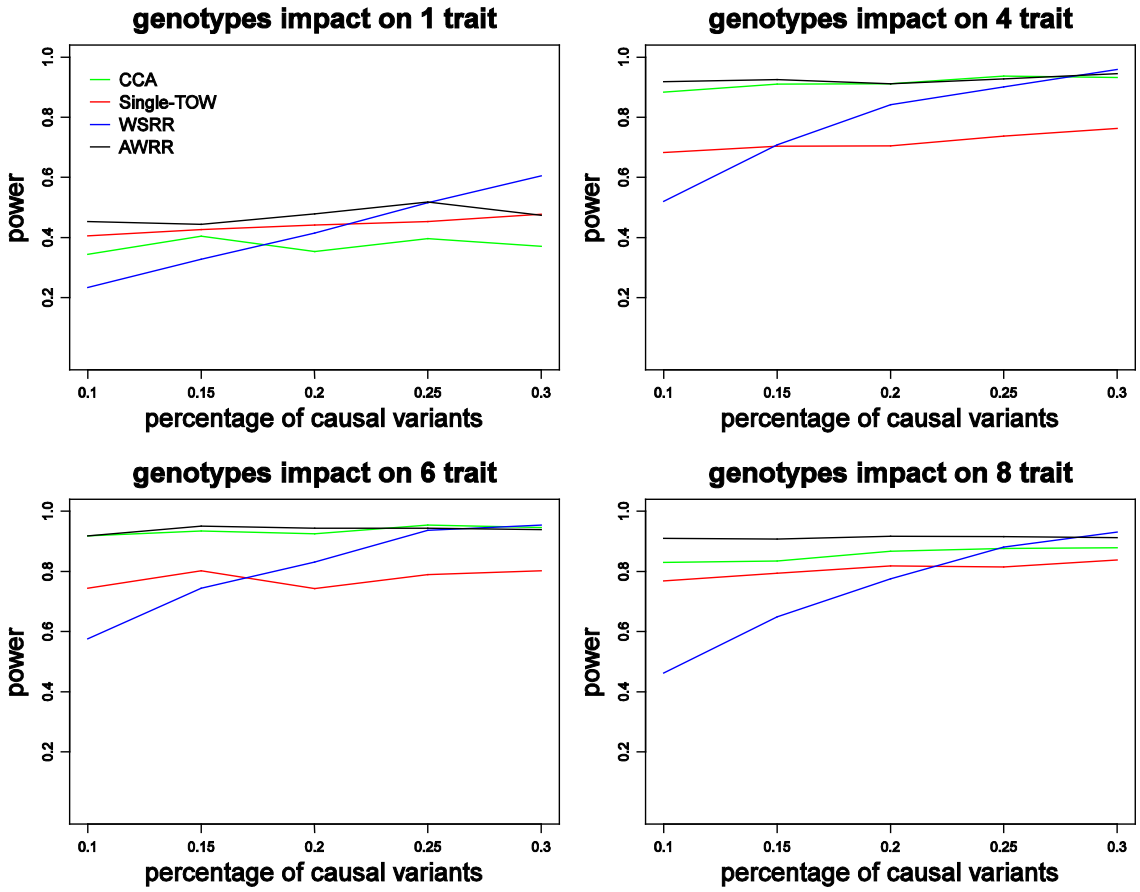


Figure 1.3. Power comparisons of the four tests (WSRR, AWRR, CCA and Single-TOW) for the power as a function of the percentage of causal variants for quantitative traits under variance model 1. The sample size is 1000 and $\rho = 0.5$, and the total heritability is 0.03. All causal variants are risk variants. The total number of traits is 10. This set of simulations is based on Sgene1.



2 Chapter 2

Joint Analysis of Multiple Traits Using "Optimal" Maximum Heritability Test

Abstract: The joint analysis of multiple traits has recently become popular since it can increase statistical power to detect genetic variants and there is increasing evidence showing that pleiotropy is a widespread phenomenon in complex diseases. Currently, most of existing methods use all of the traits for testing the association between multiple traits and a single variant. However, those methods for association studies may lose power in the presence of a large number of noise traits. In this paper, we propose an "optimal" maximum heritability test (MHT-O) to test the association between multiple traits and a single variant. MHT-O includes a procedure of deleting traits that have weak or no association with the variant. Using extensive simulation studies, we compare the performance of MHT-O with MHT, Trait-based Association Test uses Extended Simes procedure (TATES), SUM_SCORE and MANOVA. Our results show that, in all of the simulation scenarios, MHT-O is either the most powerful test or comparable to the most powerful test among the five tests we compared.

Introduction

Increasing evidence shows that pleiotropy, the effect of one variant on multiple traits, is a widespread phenomenon in complex diseases [Sivakumaran et al., 2011]. Furthermore, in genetic association studies of complex diseases, multiple related traits are usually measured. For example, hyperuricemia is usually present in patients with gout [Yang et al., 2010]; coronary heart disease is predicted by cytokine interleukin-6, C-reactive protein, interleukin-1, tumor necrosis factor-α and fibrinogen [Yudkin et al., 2000; Rifai and Ridker, 2002]; and neuropsychiatric disorders depend on a range of overlapping clinical characteristics [O'Reilly et al., 2012]. Although most published genome-wide association studies (GWASs) analyze each of the related traits separately, joint analysis of multiple traits may increase statistical power to detect genetic variants [Yang and Wang, 2012; Solovieff et al., 2013; Stephens, 2013; Zhou and Stephens, 2014]. Thus, joint analysis of multiple traits has recently become popular.

Several statistical methods have been developed for joint analysis of multiple traits. These methods can be roughly divided into three groups: combining the univariate analysis results, regression methods, and dimension reduction methods. For combining univariate analysis results, one first conducts the univariate test by performing an association test for each trait individually and then combines the univariate test statistics or combines the p-values of the univariate tests [O'Brien, 1984; Yang et al., 2010; van der Sluis et al., 2013; Kim et al., 2015]. Regression methods include mixed effect models [Korte et al., 2012; Zhou and Stephens, 2014; Casale et al., 2015], generalized estimating equation (GEE) methods [Zeger and Liang, 1986; Zhang et al., 2014], and reverse regression methods

[O'Reilly et al., 2012; Yan et al., 2013]. Mixed effect models can account for relatedness, population structure, and polygenic background effect, but it is computationally challenging. The GEE methods, based on a marginal regression model, allow the variant having different effect sizes and effect directions on different traits. These methods can also accommodate covariates and different types of traits. Reverse regression methods take genotypes as the response variable and multiple traits as independent predictors, therefore, reverse regression models do not need to know the complex distributions of traits and can be applied to a large number of mixed types of traits. Dimension reduction methods include canonical correlation analysis (CCA) [Tang and Ferreira, 2012], principal components of traits (PCT) [Aschard et al., 2014], and principal components of heritability (PCH) [Ott and Rabinowitz, 1999; Lange et al., 2004; Klei et al., 2008; Zhou et al., 2015]. CCA is to seek a linear combination of multiple variants and a linear combination of multiple traits such that the correlation between the two linear combinations reaches its maximum. The PCT methods are usually based on the first PC or first few PCs of the traits [Feng et al., 2007; Klei et al., 2008]. However, as Aschard et al. [2014] showed that testing only the first few PCs often has low power, whereas combining signals across all PCs can have greater power. Nevertheless, it is not clear how many PCs are needed, and how robust these methods are when there exists noise traits. PCH is to find a linear combination of multiple traits such that this linear combination has the maximum heritability.

In this article, we first propose a maximum heritability test (MHT). Based on MHT, we develop an "optimal" maximum heritability test (MHT-O) to test the association between multiple traits and a single variant. In each step of MHT-O, we delete one trait that has the weakest association with the variant. Then, we find the optimal number of traits and use MHT to test the association between the optimal number of traits and the variant. Using extensive simulation studies, we compare the performance of MHT-O with MHT, Trait-based Association Test uses Extended Simes procedure (TATES) [van der Sluis et al., 2013], SUM_SCORE and MANOVA [Yang and Wang, 2012]. Our results show that, in all of the simulation scenarios, MHT-O is either the most powerful test or comparable to the most powerful test among the five tests we compared.

Methods

We consider a sample with n unrelated individuals. Each individual has K (potentially correlated) traits and has been genotyped at one variant. Let $Y=(Y_1,\ldots,Y_K)^T$ denote the random vector of K traits and X denote the random variable of the genotype score at a variant. Let $y_i=(y_{i1},\ldots,y_{iK})^T$ denote the values of K traits and x_i denote the genotype score of the i^{th} individual, where x_i is the number of minor alleles that the i^{th} individual has at the variant. We can consider that $y_1,\ldots,y_n$ is a random sample from Y and $x_1,\ldots,x_n$ is a random sample from X.

Now, let us consider linear models

$$Y_k=\alpha_k+\beta_k X+\varepsilon_k \ (k=1,\ldots,K). \tag{1}$$

We partition the total phenotypic covariance of Y as $V_P=V_G+V_R$ [Falconer and Mackay, 1996]; $V_G=\mathrm{var}[\beta_1 X,\ldots,\beta_K X]=\mathrm{var}(X)\beta\beta^T$ is the genetic variance due to the genotype scores X, where $\beta=(\beta_1,\ldots,\beta_K)^T$; $V_R=\mathrm{var}[\varepsilon_1,\ldots,\varepsilon_K]$ is the residual covariance after removing the genetic effect. $\mathrm{var}(X)$ can be estimated by $\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$, $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$. β and V_R can be estimated from the linear models

$$y_{ik}=\alpha_k+\beta_k x_i+\varepsilon_{ik} \ (k=1,\ldots,K;\, i=1,\ldots,n).$$

β_k is estimated by the least square estimator. Let r_{ik} denote the estimates of residuals ε_{ik}. Then, the $(\mathrm{j},\mathrm{k})^{th}$ element of V_R is estimated by $\frac{1}{n}\sum_{i=1}^{n}r_{ij}r_{ik}$.

Let us consider a linear combination of Y, $w^TY=\sum_{k=1}^{K}w_kY_k$, where $w=(w_1,\ldots,w_K)^T$. The heritability of w^TY can be written as

$$h_w^2=\frac{w^TV_Gw}{w^TV_Pw}.$$

If we define $W=V_P^{1/2}w$, we can write h_w^2 as

$$h_w^2=\frac{W^TV_P^{-\frac{1}{2}}V_GV_P^{-\frac{1}{2}}W}{W^TW}=\frac{W^TVW}{W^TW},$$

where $V = V_P^{-\frac{1}{2}} V_G V_P^{-\frac{1}{2}}$. The heritability of $w^T Y$ depends on w and we can find a linear combination of $w^T Y$ that has the largest heritability among all linear combinations of Y. We define the maximum heritability as the test statistic to test the association between these K traits and the variant. We denote this test as maximum heritability test (MHT). The MHT statistic can be written as

$$T_{MHT} = \max_w h_w^2 = \lambda_{max}\left(V_G V_P^{-1}\right) = \mathrm{var}(X)\lambda_{max}\left(\beta\beta^T V_P^{-1}\right) = \mathrm{var}(X)\beta^T V_P^{-1}\beta,$$

where $\lambda_{max}(A)$ denotes the largest eigenvalue of matrix A.

However, the test statistic T_{MHT} may lose power in the presence of a large number of noise traits. Therefore, we propose an "optimal" maximum heritability test (MHT-O) to test the association between multiple traits and the variant. MHT-O includes a procedure of deleting traits that have weak or no association with the variant. It has the following steps:

Step 1. Given traits $Y = (Y_1, \ldots, Y_K)$, initialize $r = K$ and $Y^{(r)} = Y$. Denote $T_{MHT,r}$ as T_{MHT} based on $Y^{(r)}$.

Step 2. Denote $T_{MHT,r}^{-i}$ as T_{MHT} based on $Y^{(r)}$ with the i^{th} trait deleted for $i = 1, \ldots, r$; denote $I = \arg\max_i T_{MHT,r}^{-i}$ and $T_{MHT,r-1} = T_{MHT,r}^{-I}$. Let $Y^{(r-1)}$ denote $Y^{(r)}$ with the I^{th} trait deleted and update $r = r - 1$.

Step 3. Repeat step 2 until $r = 1$.

Denote p_r as the p-value of $T_{MHT,r}$. The test statistic of MHT-O is defined as

$T_{MHT-O} = \min_{1 \le r \le K} p_r$.

We use a permutation test to evaluate the p-value of T_{MHT-O}. Intuitively, two layers of permutations are needed to estimate p_r and the overall p-value for the test statistic T_{MHT-O}. Ge et al. [Ge et al., 2003]proposed that one layer of permutation can be used to estimate these p-values. We use the permutation procedure of Ge et al. to estimate p_r and the overall p-value for the test statistic T_{MHT-O}. In details, we randomly shuffle the genotypes in each permutation. Suppose we perform B times of permutations. Let $T_{MHT,r}^{(b)}$ denote the value of $T_{MHT,r}$ based on the b^{th} permuted data, where $b = 0$ represents the original data. Then, we transfer $T_{MHT,r}^{(b)}$ to $p_r^{(b)}$ by

$$p_r^{(b)} = \frac{\#\left\{d : T_{MHT,r}^{(d)} > T_{MHT,r}^{(b)} \text{ for } d = 0,1,\ldots,B\right\}}{B}.$$

Let $p^{(b)} = \min_{1 \le r \le K} p_r^{(b)}$, Then, the p-value of T_{MHT-O} is given by

$$\frac{\#\{\text{b} : p^{(b)} < p^{(0)} \text{ for } b = 1,2,\ldots,B\}}{B}.$$

The R code of MHT-O is available at Shuanglin Zhang's homepage http://www.math.mtu.edu/~shuzhang/software.html.

Comparison of Methods

We compare our proposed method with MHT, TATES [van der Sluis et al., 2013], MANOVA [Yang and Wang, 2012], and SUM_SCORE. TATES combines p-values obtained in a standard univariate GAWS to acquire one trait-based p-value, while correcting for correlations between components. SUM_SCORE performs an association test for each trait individually to obtain the univariate score test statistic for each trait. Then, the test statistic of SUM_SCORE is the summation of the univariate score test statistics. We use asymptotic distributions to evaluate the p-values of SUM_SCORE, TATES and MANOVA.

Simulation Studies

To evaluate the type I error rates and powers of MHT and MHT-O, we generate genotypes according to minor allele frequency (MAF) and assume Hardy Weinberg equilibrium. Then, we generate K traits by the factor model [van der Sluis et al., 2013; Aschard et al., 2014]

$$y = \lambda x + c\gamma f + \sqrt{1-c^2} \times \varepsilon, \qquad (1)$$

where $y = (y_1,\ldots,y_K)^T$; x is the genotype score at the variant of interest; $\lambda = (\lambda_1,\ldots,\lambda_K)$ is the vector of effect sizes of the genetic variant on the K traits; $f = (f_1,\ldots,f_R)^T \sim MVN(0,\Sigma)$, $\Sigma = (1-\rho)I + \rho A$, A is a matrix with elements of 1, I is the identity matrix, and ρ is the correlation between factors; γ is a K by R matrix; c is a constant number; and $\varepsilon = (\varepsilon_1,\ldots,\varepsilon_K)^T$ is a vector of residuals, and $\varepsilon_1,\ldots,\varepsilon_K$ are independent, and $\varepsilon_k \sim N(0,1)$ for $k = 1,\ldots,K$.

Based on equation (1), we consider five models:

Model 1: There is only one factor and genotypes impact on all traits with the same effect size. That is, $R = 1$, $\lambda = (\beta,\ldots,\beta)^T$, and $\gamma = (1,\ldots,1)^T$.

Model 2: There are five factors and genotypes impact on one factor. That is, $R = 5$, $\lambda = (0,\ldots,0,\beta,\ldots,\beta)^T$, and $\gamma = diag(D_1,D_2,D_3,D_4,D_5)$, where $D_i = (1,\ldots,1)^T$ for $i = 1,\ldots,5$.

Model 3: There are two factors and genotypes impact on one factor. That is, $R = 2$, $\lambda = (0,\ldots,0,\beta,\ldots,\beta)^T$, and $\gamma = diag(D_1,D_2)$, where $D_i = (1,\ldots,1)^T$ for $i = 1,2$.

Model 4: There are five factors and genotypes impact on one trait. That is, $R = 5$, $\lambda = (0,\ldots,0,\beta)^T$, and $\gamma = diag(D_1,D_2,D_3,D_4,D_5)$, where $D_i = (1,\ldots,1)^T$ for $i = 1,\ldots,5$.

Model 5: There is only one factor and genotypes impact on one trait. That is, $R = 1$, $\lambda = (0,\ldots,0,\beta)^T$, and $\gamma = (1,\ldots,1)^T$.

To evaluate type I error rates of MHT and MHT-O, we let $\beta = 0$. To evaluate powers, we let $\beta > 0$. In the simulation studies for evaluation of type I error rates and powers, we set MAF = 0.3 and $\rho = 0.2$.

Simulation Results

To evaluate the type I error rates of the two proposed tests (MHT and MHT-O), we consider 20 quantitative traits. We also consider different sample sizes, different significance levels, and different models. In each simulation scenario, the p-values of MHT and MHT-O are estimated by 1,000 permutations and the type I error rates of the two tests are evaluated using 10,000 replicated samples. For 10,000 replicated samples, the 95% confidence intervals (CIs) for estimated type I error rates of nominal levels 0.05 and 0.01 are (0.046, 0.054) and (0.008, 0.012), respectively (see end of this chapter). The estimated type I error rates of the two tests are summarized in Table 2.1. From this table, we can see that 58 out of 60 (greater than 95%) estimated type I error rates are within the 95% CIs and the two estimated type I error rates (0.05415 and 0.0126) not within the 95% CIs are very close to the bound of the corresponding 95% CI, which indicates that the two tests are all valid.

For power comparisons, we consider different values of the effect size, different models, and different numbers of traits. Sample size is 1,000 for all the cases. In each of the simulation scenarios, the p-values of MHT and MHT-O are estimated using 1,000 permutations and the p-values of SUM_SCORE, TATES and MANOVA are estimated using their asymptotic distributions. The powers of all of the five tests are evaluated using 500 replicated samples at a significance level of 0.05.

Figure 2.1 gives the power comparisons of the five tests (SUM_SCORE, TATES, MHT, MHT-O and MANOVA) for the power as a function of the effect size based on the five models for 20 traits. This figure shows that (1) MHT-O is either the most powerful one (genotypes directly impact on a single trait: models 4-5) or comparable to the most powerful one (genotypes directly impact on all or a portion of the traits: models 1-3) among the five tests; (2) MHT and MANOVA have very similar powers; (3) MHT and MANOVA are much less powerful than other methods when genotypes directly impact on only a portion of the traits (models 2-3); (4) TATES is much less powerful than other methods when genotypes directly impact on all the traits (model 1); and (5) SUM_SCORE is much less powerful than other methods when genotypes directly impact on a single trait (models 4-5).

Power comparisons of the five tests for 30 and 40 traits are given in Figures 2.2 and 2.3, respectively. The patterns of power comparisons for 30 and 40 traits (Figures 2.2 and 2.3) are similar to that for 20 traits (Figure 2.1). We also give power comparisons of the five tests using a significance level of 5×10^{-8} with 10^8 permutations and 500 replicates for 20 traits under model 1 (Figure B.2.1). Figure B.2.1 shows that the patterns of the power comparisons using significance level 5×10^{-8} are similar to that using a significance level of 0.05 in Figure 2.1 (model 1). In summary, MHT-O is either the most powerful test or comparable to the most powerful test among all the tests we compared. Therefore, our MHT-O is a robust test to a variety of models.

Discussion

We propose MHT-O to perform joint analysis of multiple traits in association studies based on the following reasons: (1) multiple related traits are usually measured in genetic association studies of complex diseases; (2) there is increasing evidence showing that pleiotropy is a widespread phenomenon in complex diseases; and (3) the power of existing methods decreases in the presence of non-associated traits. The proposed MHT-O includes a procedure of deleting traits that have weak or no association with the variant. Therefore, it can be robust to the existence and the number of non-associated traits. By deleting one trait that has the weakest association with the variant in each step, MHT-O can maintain high power in the presence of a large number of non-associated traits. This feature is essentially important when there exist a large number of correlated traits but there are no guidelines to select relevant traits. Our results show that MHT-O has correct type I error rates and is either the most powerful test or comparable to the most powerful test among the five tests we compared. No other methods in the simulation studies show consistent good performance.

Due to the allelic heterogeneity and the extreme rarity of individual variants in rare variant association studies, the variant-by-variant methods for common variant association studies may not be optimal [Li and Leal, 2008]. It has been shown by recent studies that complex diseases are caused by both common and rare variants [Pritchard, 2001; Pritchard and Cox, 2002; Walsh and King, 2007; Bodmer and Bonilla, 2008; Stratton and Rahman, 2008; Kang et al., 2010; Teer and Mullikin, 2010]. Statistical methods including burden tests [Morgenthaler and Thilly, 2007; Li and Leal, 2008; Madsen and Browning, 2009; Price et al., 2010; Zawistowski et al., 2010], quadratic tests [Neale et al., 2011; Wu et al., 2011; Sha et al., 2012], and combined tests [Lee et al., 2012; Derkach et al., 2013; Sha and Zhang, 2014] have been developed for rare variant association studies with a single trait. Currently, there are limited researches on rare variant association studies for joint analysis of multiple traits [Casale et al., 2015; Wang et al., 2015]. MHT-O can be extended to rare variant association studies by extending equation (1) to include multiple variants. MHT-O can also be extended to family-based studies by extending equation (1) to mixed linear model. However, the performance of MHT-O in rare variant association studies and in family-based association studies needs further investigation.

The fact that population stratification can seriously confound association results has been long recognized in association studies based on unrelated individuals [Knowler et al., 1988; Lander and Schork, 1994]. Several methods to control for population stratification have been developed for association studies based on unrelated individuals. These methods include principal component (PC) approach [Zhu et al., 2002; Chen et al., 2003; Zhang et al., 2003; Price et al., 2006; Bauchet et al., 2007], genomic control (GC) approach [Devlin and Roeder, 1999; Devlin et al., 2001; Reich and Goldstein, 2001], and mixed linear model (MLM) approach [Kang et al., 2010; Zhang et al., 2010]. Like most association tests based on unrelated individuals, MHT-O subjects to bias due to population stratification. To make MHT-O robust to population stratification, we can use the PC approach. Let

$P_i = (p_{i1}, \ldots, p_{iL})^T$ denote the first L PCs of the genotypes at a set of genomic markers for the i^{th} individual. Let y_{ik}^* and x_i^* denote the residuals of the regressions $y_{ik} = \alpha_{0k} + \alpha_k^T P_i + \varepsilon_{ik}$ and the residuals of the regression $x_i = \alpha_0 + \alpha^T P_i + \varepsilon_i$, respectively. Using y_{ik}^* and x_i^* to replace y_{ik} and x_i, we can make MHT-O robust to population stratification. However, the performance of using the PC approach to control for population stratification in MHT-O needs further investigations.

Confidence Interval

Let p denote the p-value of the test and denote a random variable

$$\xi = \begin{cases} 1, & p \le \alpha \\ 0, & p > \alpha \end{cases},$$

where α is the significance level. Then, $\Pr(\xi = 1) = \alpha$ and $\Pr(\xi = 0) = 1 - \alpha$ because p follows a uniform distribution between 0 and 1 under the null hypothesis. Suppose there are R replicates. Let ξ_i denote the value of ξ for the i^{th} replicate, where $i = 1, \ldots, R$. Then, the estimated type I error rate is given by $\bar{\xi} = \frac{1}{R}\sum_{i=1}^{R} \xi_i$ that asymptotically follows a normal distribution $N\left(\alpha, \frac{\alpha(1-\alpha)}{R}\right)$. Thus,

$$\Pr\left(\left|\frac{\bar{\xi} - \alpha}{\sqrt{\alpha(1-\alpha)/R}}\right| \le 1.96\right) = \Pr\left(\alpha - 1.96\sqrt{\alpha(1-\alpha)/R} \le \bar{\xi} \le \alpha + 1.96\sqrt{\alpha(1-\alpha)/R}\right) = 0.95 .$$

We define $\left(\alpha - 1.96\sqrt{\alpha(1-\alpha)/R}, \alpha + 1.96\sqrt{\alpha(1-\alpha)/R}\right)$ as the 95% confidence interval for the estimated type I error rate for the nominal level α.

Tables and Figures

Table 2.1. The estimated type I error rates of MHT and MHT-O. 10,000 replicates are used.

		Sample size			
			500	1000	2000
Model 1	$\alpha = 0.05$	MHT-O	0.05415	0.0494	0.04875
		MHT	0.05235	0.05005	0.0501
	$\alpha = 0.01$	MHT-O	0.01035	0.012	0.0091
		MHT	0.00985	0.01195	0.01105
Model 2	$\alpha = 0.05$	MHT-O	0.0499	0.0515	0.0526
		MHT	0.04815	0.05175	0.05285
	$\alpha = 0.01$	MHT-O	0.01045	0.01175	0.01135
		MHT	0.0117	0.0118	0.0126
Model 3	$\alpha = 0.05$	MHT-O	0.05015	0.0517	0.05315
		MHT	0.04875	0.0507	0.0529
	$\alpha = 0.01$	MHT-O	0.00995	0.0109	0.012
		MHT	0.0104	0.01035	0.012
Model 4	$\alpha = 0.05$	MHT-O	0.04815	0.0516	0.05255
		MHT	0.04875	0.05275	0.0507
	$\alpha = 0.01$	MHT-O	0.00975	0.0118	0.00975
		MHT	0.00855	0.012	0.01
Model 5	$\alpha = 0.05$	MHT-O	0.04865	0.0499	0.04975
		MHT	0.05095	0.05195	0.04755
	$\alpha = 0.01$	MHT-O	0.012	0.0119	0.00915
		MHT	0.01075	0.01115	0.0096

Figure 2.1. Power comparisons of the five tests (SUM_SCORE, TATES, MHT, MHT-O and MANOVA) for the power as a function of the effect size. Sample size is 1000. Total number of traits is 20.

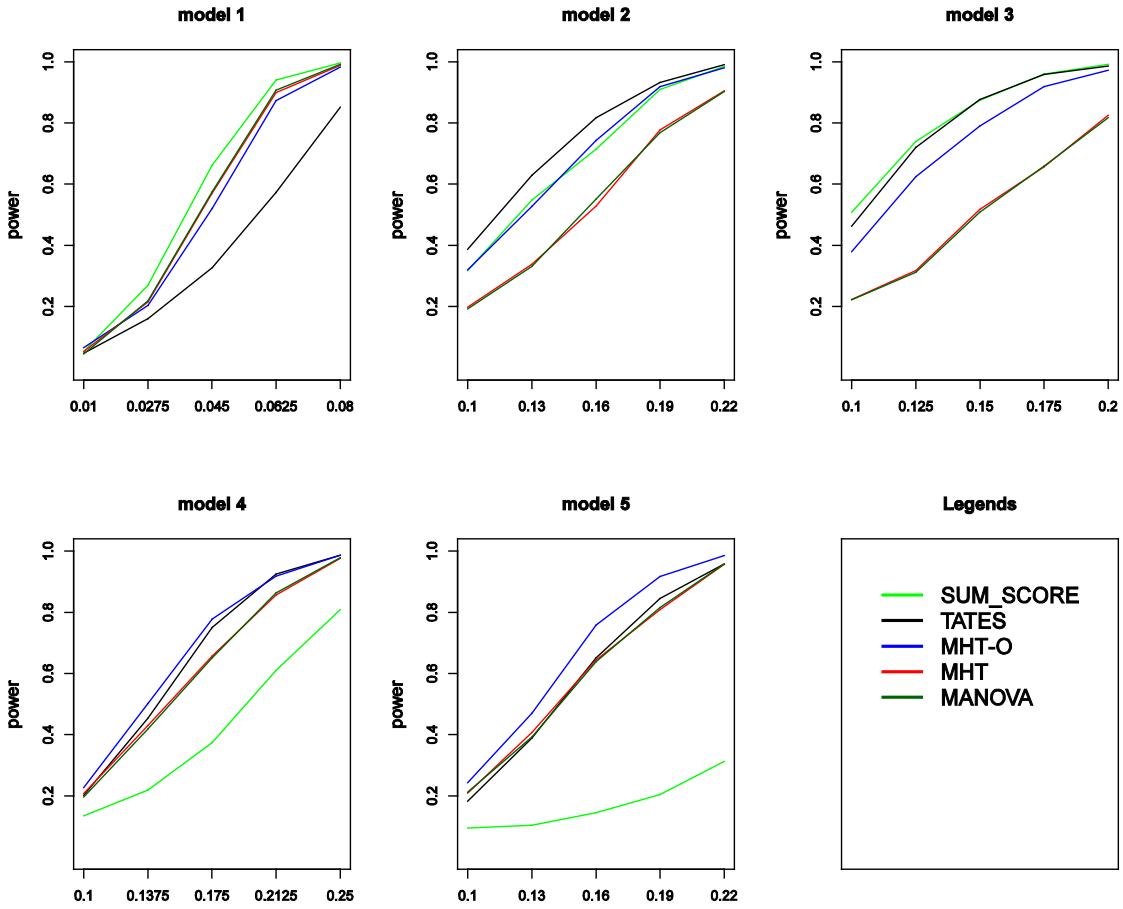


Figure 2.2. Power comparisons of the five tests (SUM_ SCORE, TATES, MHT, MHT-O and MANOVA) for the power as a function of the effect size. Sample size is 1000. Total number of traits is 30.

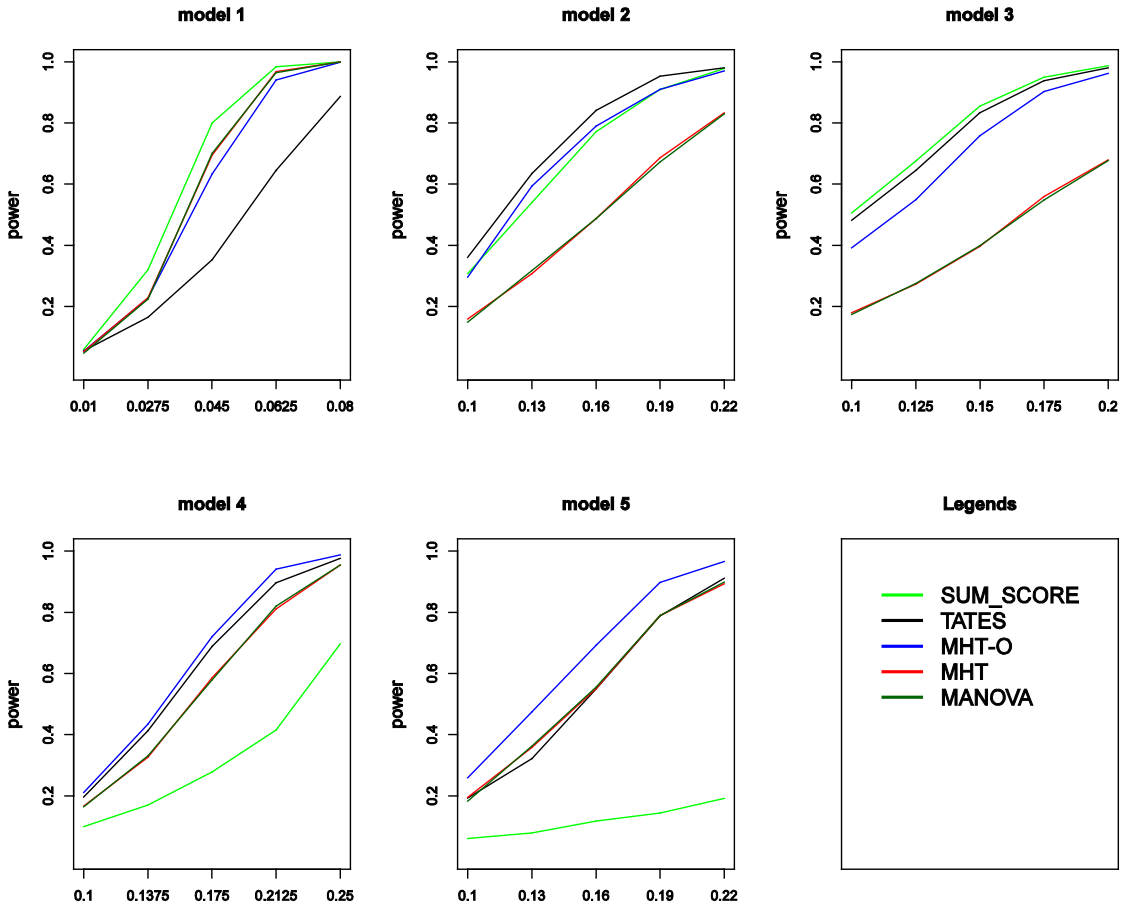
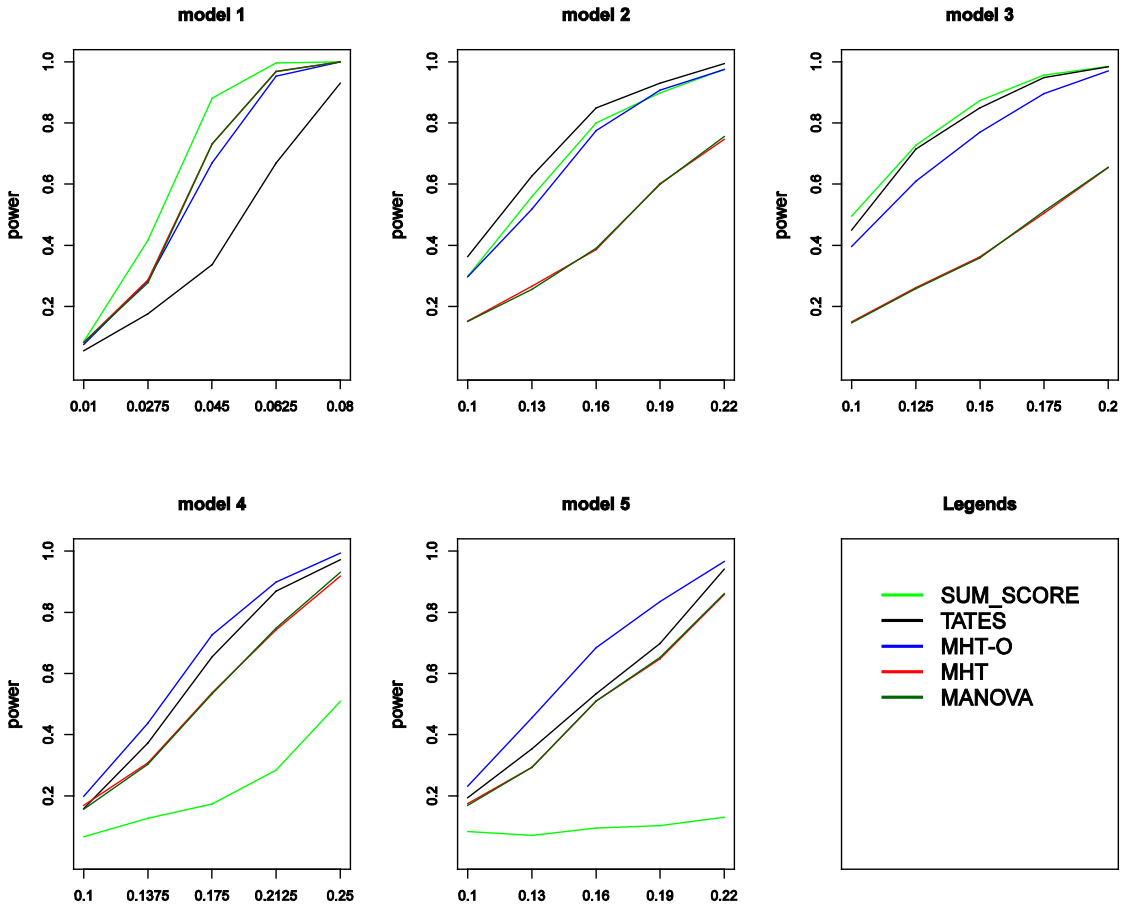


Figure 2.3. Power comparisons of the five tests (SUM_SCORE, TATES, MHT, MHT-O and MANOVA) for the power as a function of the effect size. Sample size is 1000. Total number of traits is 40.



Chapter 3

Testing an optimally weighted combination of common and/or rare variants with multiple traits

Joint analysis of multiple traits has recently become popular since it can increase statistical power to detect genetic variants and there is increasing evidence showing that pleiotropy is a widespread phenomenon in complex diseases. Currently, most of existing methods test the association between multiple traits and a single common variant. However, the variant-by-variant methods for common variant association studies may not be optimal for rare variant association studies due to the allelic heterogeneity as well as the extreme rarity of individual variants. In this article, we developed a statistical method by testing an optimally weighted combination of variants with multiple traits (TOWmuT) to test the association between multiple traits and a weighted combination of variants (rare and/or common) in a genomic region. TOWmuT is robust to the directions of effects of causal variants and is applicable to different types of traits. Using extensive simulation studies, we compared the performance of TOWmuT with the following five existing methods: gene association with multiple traits (GAMuT), multiple sequence kernel association test (MSKAT), adaptive weighting reverse regression (AWRR), single-TOW, and MANOVA. Our results showed that, in all of the simulation scenarios, TOWmuT has correct type I error rates and is consistently more powerful than the other five tests. We also illustrated the usefulness of TOWmuT by analyzing a whole-genome genotyping data from a lung function study.

Introduction

Many large cohort studies collected many correlated traits that can reflect underlying mechanism of complex diseases. For example, the UK10K cohort study collected 64 correlated phenotypic traits [The UK10K Consortium et al., 2015]. Usually complex diseases are characterized by multiple endophenotypes. For example, hypertension can be characterized by systolic and diastolic blood pressure [Newton-Cheh et al., 2009]; metabolic syndrome is evaluated by four component traits: high-density lipoprotein (HDL) cholesterol, plasma glucose and Type 2 diabetes, abdominal obesity, and diastolic blood pressure [Zabaneh and Balding, 2010]; and schizophrenia can be diagnosed by eight neurocognitive domains [Gur et al., 2007]. Multiple correlated traits can be influenced by a gene simultaneously. Therefore, by joint analysis of multiple traits, we can not only gain more statistical power to detect pleiotropic variants [Yang and Wang, 2012; Solovieff et al., 2013; Stephens, 2013; Zhou and Stephens, 2014; Zhu et al., 2015a; Liang et al., 2016; Wang et al., 2016a; Wang et al., 2016b], but also can be important to understand the genetic architecture of the disease of interest [Aschard et al., 2014].

Several statistical methods have been developed for testing the association between multiple traits and a single common variant. These methods can be roughly divided into

three groups: dimension reduction methods [Klei et al., 2008; Ferreira and Purcell, 2009; Aschard et al., 2014; Wang et al., 2016a], regression methods [Korte et al., 2012; O'Reilly et al., 2012; Zhang et al., 2014], and combining test statistics from univariate analysis [O'Brien, 1984; Yang et al., 2010; van der Sluis et al., 2013; Kim et al., 2015; Zhu et al., 2015b; Liang et al., 2016]. However, due to the allelic heterogeneity as well as the extreme rarity of rare variants [Li and Leal, 2008], the variant-by-variant methods for common variant association studies may not be optimal for rare variant association studies. Recent studies show that complex diseases are caused by both common and rare variants [Pritchard, 2001; Pritchard and Cox, 2002; Walsh and King, 2007; Bodmer and Bonilla, 2008; Stratton and Rahman, 2008; Kang et al., 2010; Teer and Mullikin, 2010]. Next-generation sequencing technology allows sequencing of the whole genome of large groups of individuals, and thus makes rare variant association studies feasible [Andres et al., 2007; Metzker, 2010]. Recently, statistical methods for rare variant association studies with a single trait have been developed by summarizing genotype information from multiple variants. These methods include burden tests [Morgenthaler and Thilly, 2007; Li and Leal, 2008; Madsen and Browning, 2009; Price et al., 2010; Zawistowski et al., 2010], quadratic tests [Neale et al., 2011; Wu et al., 2011; Sha et al., 2012; Yang et al., 2017], and combined tests [Derkach et al., 2013; Lee et al., 2013; Sha and Zhang, 2014; Greco et al., 2015].

As we pointed out above, it is essential to develop statistical methods to test the association between multiple traits and multiple variants (common and/or rare variants). Very recently, a few statistical methods for this purpose are appeared [Casale et al., 2015; Wang et al., 2015; Broadaway et al., 2016; Sun et al., 2016; Wang et al., 2016b; Wu and Pankow, 2016]. Casale et al. [2015] proposed a set-based association test based on the linear mixed-model. This method enables jointly analyzing multiple correlated traits in rare variant association studies while accounting for population structure and relatedness. Wang et al. [2015] proposed a multivariate functional linear model approach to test association between multiple traits and rare variants in a genomic region. In this approach, the genetic effects of variants are treated as smooth functions of genomic positions of these variants. Gene association with multiple traits (GAMuT) proposed by Broadaway et al. [2016] provide a nonparametric test of independence between a set of traits and a set of genetic variants. This method compares the similarities of multiple traits with the similarities of genotypes at variants in a genomic region. Multivariate Rare-Variant Association Test (MURAT) proposed by Sun et al. [2016] tests association between multiple correlated quantitative traits and a set of rare variants based on a linear mixed model. This method assumes that the effects of the variants follow a multivariate normal distribution with a zero mean and a specific covariance structure. Wu and Pankow [2016] extended the commonly used sequence kernel association test (SKAT) [Wu et al., 2011] for a single trait to multiple traits and proposed multiple sequence kernel association test (MSKAT). Wang et al. [2016b] proposed an adaptive weighting reverse regression (AWRR) method. This method uses the score test based on the reverse regression, in which the summation of

adaptively weighted genotypes is treated as the response variable and multiple traits are treated as independent variables.

In this article, we developed a new statistical method by testing an optimally weighted combination of variants with multiple traits (TOWmuT) to test the association between multiple traits and a weighted combination of variants (rare and/or common) in a genomic region. TOWmuT is based on the score test under a linear model, in which the weighted combination of variants is treated as the response variable and multiple traits including covariates are treated as independent variables. The statistic of TOWmuT is the maximum of the score test statistic over weights. The weights at which the score test statistic reaches its maximum are called the optimal weights. TOWmuT is applicable to different types of traits and can include covariates. Using extensive simulation studies, we compared the performance of TOWmuT with single-TOW [Sha et al., 2012], GAMuT [Broadaway et al., 2016], MSKAT [Wu and Pankow, 2016], AWRR [Wang et al., 2016b] and MANOVA [Yang and Wang, 2012]. Our results showed that, in all the simulation scenarios, TOWmuT is either the most powerful test or comparable to the most powerful test among the six tests. We also illustrated the usefulness of TOWmuT by analyzing a real whole-genome genotyping data from a lung function study.

Methods

We consider a sample with n unrelated individuals. Each individual has K potentially correlated quantitative or qualitative traits (1 for cases and 0 for controls for a qualitative trait) and has been genotyped at M variants in a genomic region. Let y_{ik}^* denote the k^{th} trait value of the i^{th} individual and x_{im}^* denote the genotype score of the i^{th} individual at the m^{th} variant, where x_{im}^* is the number of minor alleles that the i^{th} individual carries at the m^{th} variant. We first centralize y_{ik}^* and x_{im}^* as $y_{ik} = y_{ik}^* - \bar{y}_k$ and $x_{im} = x_{im}^* - \bar{x}_m$, where $\bar{y}_k = \frac{1}{n}\sum_{i=1}^{n} y_{ik}^*$ and $\bar{x}_m = \frac{1}{n}\sum_{i=1}^{n} x_{im}^*$. Let $Y_i = \left(y_{i1},\ldots,y_{iK}\right)^T$, $X_i = \left(x_{i1},\ldots,x_{iM}\right)^T$, $Y = \left(Y_1,\ldots,Y_n\right)^T$, and $X = \left(X_1,\ldots,X_n\right)^T$. For the i^{th} individual, we consider a linear combination of the variants $x_i = \sum_{m=1}^{M} w_m x_{im}$, where $w = \left(w_1,\ldots,w_M\right)^T$ are weights and their values will be decided later.

Without covariates

We first describe our method without covariates. Consider the linear model

$$x_i = \beta_1 y_{i1} + \cdots + \beta_K y_{iK} + \varepsilon_i. \tag{1}$$

The score test statistic to test the null hypothesis $H_0: \beta_1 = \cdots = \beta_K = 0$ is given by

$$T_{score} = U^T V^{-1} U / \sigma^2, \tag{2}$$

where $U = \sum_{i=1}^{n} x_i Y_i = Y^T X w$, $V = \sum_{i=1}^{n} Y_i Y_i^T = Y^T Y$, and $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 = \frac{1}{n} w^T X^T X w$. We use $A = diag\left(\frac{1}{n} X^T X\right)$ to replace $\frac{1}{n} X^T X$. Then σ^2 becomes $\sigma_0^2 = w^T A w$ and T_{score} becomes $T_{score}^0(w) = \frac{w^T X^T Y \left(Y^T Y\right)^{-1} Y^T X w}{w^T A w}$. We define the test statistic of TOWmuT as

$$T_{TOWmuT} = \max_w T_{score}^0(w). \tag{3}$$

Let $W = A^{1/2} w$, then $T_{TOWmuT} = \max_W T_{score}^0(W) = \lambda_{max}\left(A^{-1/2} X^T Y \left(Y^T Y\right)^{-1} Y^T X A^{-1/2}\right)$, where $\lambda_{\max}$ indicates the largest eigenvalue of a matrix. Let W^0 denote the eigenvector of $A^{-1/2} X^T Y \left(Y^T Y\right)^{-1} Y^T X A^{-1/2}$ corresponding to the largest eigenvalue, then $w^0 = A^{-1/2} W^0$ is the optimal weights. Actually, we do not need to calculate w^0 in order to calculate T_{TOWmuT}. If we let $C = X A^{-1} X^T$, then

$$T_{TOWmuT} = \lambda_{max}\left(A^{-1/2}X^TY\left(Y^TY\right)^{-1}Y^TXA^{-1/2}\right) = \lambda_{max}\left(\left(Y^TY\right)^{-1}Y^TCY\right). \quad (4)$$

We use a permutation test to evaluate the p-value of T_{TOWmuT}. In details, we randomly shuffle the traits in each permutation. Note that C and $\left(Y^TY\right)^{-1}$ do not change in each permutation. Suppose that we perform B times of permutations. Let $T_{TOWmuT}^{(b)}$ denote the value of T_{TOWmuT} based on the b^{th} permuted data, where $b=0$ represents the original data. Then, the p-value of T_{TOWmuT} is given by

$$\frac{\#\left\{b : T_{TOWmuT}^{(b)} \geq T_{TOWmuT}^{(0)} \text{ for } b = 1,\ldots,B\right\}}{B}. \quad (5)$$

With covariates

Assume that there are p covariates and $z_{i1},\ldots,z_{ip}$ denote the p covariates of the i^{th} individual. Consider the linear model

$$x_i = \alpha_0 + \alpha_1 z_{i1} + \cdots + \alpha_p z_{ip} + \beta_1 y_{i1} + \cdots + \beta_K y_{iK} + \varepsilon_i. \quad (6)$$

In the end of this chapter, we showed that under model (6), the score test statistic with covariates to test the null hypothesis $H_0 : \beta_1 = \mathsf{L} = \beta_K = 0$ is given by

$$T_{score}^c = \tilde{U}^T\tilde{V}\tilde{U}/\tilde{\sigma}^2, \quad (7)$$

where $\tilde{U} = \tilde{Y}^T\tilde{X}w$, $\tilde{V} = \tilde{Y}^T\tilde{Y}$, $\tilde{\sigma}^2 = \frac{1}{n}w^T\tilde{X}^T\tilde{X}w$, $\tilde{X} = (\tilde{x}_{im})$, $\tilde{Y} = (\tilde{y}_{ik})$, $\tilde{y}_{ik}$ and $\tilde{x}_{im}$ denote the residuals of y_{ik} and x_{im} under

$$y_{ik} = \alpha_{0k} + \alpha_{1k}z_{i1} + \ldots + \alpha_{pk}z_{ip} + \varepsilon_{ik} \text{ and } x_{im} = \alpha_{0m} + \alpha_{1m}z_{i1} + \ldots + \alpha_{pm}z_{ip} + \tau_{im}. \quad (8)$$

We can see the score test statistic with covariates

$$T_{score}^c = T_{score}. \quad (9)$$

That is, replacing y_{ik} and x_{im} by their residuals $\tilde{y}_{ik}$ and $\tilde{x}_{im}$ in the score test statistic without covariates T_{score}, it becomes the score test statistic with covariates T_{score}^c.

Therefore, we define TOWmuT statistic with covariates as

$$T_{TOWmuT}^c = T_{TOWmuT}. \quad (10)$$

In summary, to apply TOWmuT with covariates, we adjust both trait value y_{ik} and genotypic score x_{im} for the covariates by applying linear regressions in (8) and apply TOWmuT without covariates to the residuals $\tilde{y}_{ik}$ and $\tilde{x}_{im}$.

Comparison of Methods

We compare the performance of our method with the following methods: Multivariate Analysis of Variance (MANOVA) [Liang et al., 2016], MSKAT [Wu and Pankow, 2016], GAMuT [Broadaway et al., 2016], AWRR [Wang et al., 2016b] and single-TOW [Sha et al., 2012]. Here we briefly introduce each of those methods using the notations in the method section.

MANOVA: Consider a multivariate multiple linear regression model: $Y = X\beta + \varepsilon$, where Y denotes the $n \times K$ matrix of phenotypes; X denotes the $n \times M$ matrix of genotypes; β is a $M \times K$ matrix of coefficients; ε is the $n \times K$ matrix of random errors with each row of ε to be i.i.d. $MVN(0,\Sigma)$, where Σ is the covariance matrix of ε. To test $H_0 : \beta = 0$, the likelihood ratio test is equivalent to the Wilk's Lambda test statistic of MANOVA, that is, $-2\log\Lambda = 2\left(l\left(\hat{\beta},\hat{\Sigma}\right) - l\left(0,\hat{\Sigma}_0\right)\right) = n\log\frac{\left|\hat{\Sigma}_0\right|}{\left|\hat{\Sigma}\right|} = -n\log\left(\frac{|E|}{|E+H|}\right)$. Here Λ denote the ratio of the likelihood function under H_0 to the likelihood function under H_1 , $l(\beta,\Sigma)$ is the log-likelihood function, $H = \hat{\beta}^T\left(X^TX\right)\hat{\beta}$ and $E = Y^TY - \hat{\beta}^T\left(X^TX\right)\hat{\beta}$, where $\hat{\beta} = \left(X^TX\right)^{-1}X^TY$ is the maximum likelihood estimator (MLE) of β , and $|\ |$ denotes the determinant of a matrix. The test statistic has an asymptotic χ^2_K distribution.

MSKAT: MSKAT extends the commonly used SKAT [Wu et al., 2011] for single trait analysis to test for the joint association of rare variant set with multiple continuous traits.

GAMuT: GAMuT compares the similarity in multivariate phenotypes to the similarity in rare-variant genotypes in a genomic region by a machine-learning framework called kernel distance covariance.

AWRR: by collapsing genotypes using adaptive weights, AWRR uses the score test to test association based on the reverse regression, in which collapsed genotypes are treated as the response variable and multiple traits are treated as independent variables.

Single-TOW: Let T^k_{TOW} denote the test statistic of TOW to test the association between the k th trait and the genotypes at the variants in a genomic region. The test

statistic of single-TOW is given by $T_{single-TOW} = \min_{1 \le k \le K} p_k$, where p_k is the p-value of T_{TOW}^k for $k = 1, \ldots, K$. The p-value of $T_{single-TOW}$ is estimated using a permutation procedure.

Simulations

In our simulation studies, we use the empirical Mini-Exome genotype data provided by the genetic analysis workshop 17 (GAW17) to generate genotypes. This dataset contains genotypes of 697 unrelated individuals on 3205 genes. We choose four genes: ELAVL4 (gene1), MSH4 (gene2), PDE4B (gene3), and ADAMTS4 (gene4) with 10, 20, 30, and 40 variants, respectively. Then, we merge the four genes to form a super gene (Sgene) with 100 variants. In our simulation studies, we generate genotypes based on the genotypes of 697 individuals in the Sgene because the distribution of the minor allele frequencies (MAFs) in the Sgene can represent the distribution of MAFs in all of the 3205 genes [Sha et., 2012]. To generate a qualitative disease affection status, we use a liability threshold model based on a continuous phenotype (quantitative trait). An individual is defined as affected if the individual's phenotype is at least one standard deviation larger than the phenotypic mean. This yields a prevalence of 16% for the simulated disease in the general population. In the following, we describe how to generate a quantitative trait.

We consider that all causal variants are rare (MAF < 0.01). We randomly choose n_c rare variants as causal variants, where n_c is determined by the percentage of causal variants among rare variants. We use n_r and n_p to denote the number of risk rare variants and protective rare variants, respectively, where $n_r + n_p = n_c$. Let x_{qi}^r and x_{ji}^p denote the genotypic scores of the q^{th} risk rare variant and the j^{th} protective rare variant for the i^{th} individual, respectively. We assume that genotypes impact on L traits. Let h and h_l denote the heritability of all the n_c rare causal variants for the L traits and the l^{th} trait among the L traits, respectively. We generate L random numbers $t_1,\ldots,t_L$ from a uniform distribution between 0 and 1. Then, the heritability of l^{th} trait among the L traits is $h_l = ht_l \Big/ \sum_{l=1}^{L} t_l$. Given the heritability of the l^{th} trait h_l, we generate n_c random numbers $r_1,\ldots,r_{n_c}$ from a uniform distribution between 0 and 1. The heritability of the m^{th} causal variant for the l^{th} trait is given by $h_l^{(m)} = h_l r_m \Big/ \sum_{j=1}^{n_c} r_j$.

In our simulation studies, we consider two covariates Z_1 and Z_2, where Z_1 is a continuous covariate generated from a standard normal distribution, and Z_2 is a binary covariate taking values 0 and 1 with a probability of 0.5. We generate K traits by considering the factor model [van der Sluis et al., 2013; Aschard et al., 2014; Wang et al., 2016a]

$$y = (0.5Z_1 + 0.5Z_2)e + (\lambda_1,\ldots,\lambda_K)^T + c\gamma f + \sqrt{1-c^2} \times \varepsilon, \tag{11}$$

where $y=(y_1,\ldots,y_K)^T$; $e=(1,\ldots,1)^T$; $\lambda=(\lambda_1,\ldots,\lambda_K)$ is the vector involved genotypes; $f=(f_1,\ldots,f_R)^T \sim MVN(0,\Sigma)$, $\Sigma=(1-\rho)I+\rho A$, A is a matrix with elements of 1, I is the identity matrix, and ρ is the correlation between f_i and f_j; R is the number of factors; γ is a K by R matrix; c is a constant number; $\varepsilon=(\varepsilon_1,\ldots,\varepsilon_K)^T$ is a vector of residuals; and $\varepsilon_1,\ldots,\varepsilon_K$ are independent, $\varepsilon_k \sim N(0,1)$ for $k=1,\ldots,K$.

We consider the following six models with different number of factors and different number of traits affected by genotypes. In these models, the within-factor correlation is c^2 and the between-factor correlation is $\rho_1=\rho c^2$.

Model 1: There is only one factor and genotypes impact on 6 traits with the same effect size. This is equivalent to set $R=1$ and $\gamma=(1,\ldots,1)^T$. In details,

$$y_k=\begin{cases}0.5Z_1+0.5Z_2+\sum_{q=1}^{n_r}\beta_{kq}^r x_q^r-\sum_{j=1}^{n_p}\beta_{kj}^p x_j^p+cf_1+\sqrt{1-c^2}\times\varepsilon_k, & 1\le k\le 6\\ 0.5Z_1+0.5Z_2+cf_1+\sqrt{1-c^2}\times\varepsilon_k, & k>6\end{cases}. \quad (12)$$

Model 2: There are five factors and genotypes impact on 6 traits. We set $R=5$ and $\gamma=diag(D_1,D_2,D_3,D_4,D_5)$, where $D_i=(1,\ldots,1)^T$ for $i=1,\ldots,5$. In details,

$$y_k=\begin{cases}0.5Z_1+0.5Z_2+\sum_{q=1}^{n_r}\beta_{kq}^r x_q^r-\sum_{j=1}^{n_p}\beta_{kj}^p x_j^p+cf_{[(k-1)/2]+1}+\sqrt{1-c^2}\times\varepsilon_k, & 1\le k\le 6\\ 0.5Z_1+0.5Z_2+cf_{[(k-1)/2]+1}+\sqrt{1-c^2}\times\varepsilon_k, & k>6\end{cases}. \quad (13)$$

Model 3: There are two factors and genotypes impact on 6 traits. That is, $R=2$ and $\gamma=diag(D_1,D_2)$, where $D_i=(1,\ldots,1)^T$ for $i=1,2$. In details,

$$y_k=\begin{cases}0.5Z_1+0.5Z_2+\sum_{q=1}^{n_r}\beta_{kq}^r x_q^r-\sum_{j=1}^{n_p}\beta_{kj}^p x_j^p+cf_{[(k-1)/5]+1}+\sqrt{1-c^2}\times\varepsilon_k, & 1\le k\le 6\\ 0.5Z_1+0.5Z_2+cf_{[(k-1)/5]+1}+\sqrt{1-c^2}\times\varepsilon_k, & k>6\end{cases}. \quad (14)$$

Model 4: There are five factors and genotypes impact on one trait. That is, $R=5$ and $\gamma=diag(D_1,D_2,D_3,D_4,D_5)$, where $D_i=(1,\ldots,1)^T$ for $i=1,\ldots,5$. In details,

$$y_k = \begin{cases} 0.5Z_1 + 0.5Z_2 + \sum_{q=1}^{n_r} \beta_{kq}^r x_q^r - \sum_{j=1}^{n_p} \beta_{kj}^p x_j^p + cf_{\lfloor (k-1)/2 \rfloor + 1} + \sqrt{1-c^2} \times \varepsilon_k, & k = 1 \\ 0.5Z_1 + 0.5Z_2 + cf_{\lfloor (k-1)/2 \rfloor + 1} + \sqrt{1-c^2} \times \varepsilon_k, & k > 1 \end{cases}. \quad (15)$$

Model 5: There are only two factors and genotypes impact on one trait. That is, $R = 2$ and $\boldsymbol{\gamma} = diag(D_1, D_2)$, where $D_i = (1, \ldots, 1)^T$ for $i = 1, 2$. In details,

$$y_k = \begin{cases} 0.5Z_1 + 0.5Z_2 + \sum_{q=1}^{n_r} \beta_{kq}^r x_q^r - \sum_{j=1}^{n_p} \beta_{kj}^p x_j^p + cf_{\lfloor (k-1)/5 \rfloor + 1} + \sqrt{1-c^2} \times \varepsilon_k, & k = 1 \\ 0.5Z_1 + 0.5Z_2 + cf_{\lfloor (k-1)/5 \rfloor + 1} + \sqrt{1-c^2} \times \varepsilon_k, & k > 1 \end{cases}. \quad (16)$$

Model 6: There is K factors and genotypes impact on 6 traits. That is, $R = K$, $\boldsymbol{\gamma} = I$, and $c = 1$. In details,

$$y_k = \begin{cases} 0.5Z_1 + 0.5Z_2 + \sum_{q=1}^{n_r} \beta_{kq}^r x_q^r - \sum_{j=1}^{n_p} \beta_{kj}^p x_j^p + cf_k + \sqrt{1-c^2} \times \varepsilon_k, & 1 \le k \le 6 \\ 0.5Z_1 + 0.5Z_2 + cf_k + \sqrt{1-c^2} \times \varepsilon_k, & k > 6 \end{cases}. \quad (17)$$

Results

To evaluate the type I error rates of the proposed test TOWmuT, we set $\lambda_k = 0$ for $k = 1,\ldots,K$ in the 6 models. We consider different sample sizes, different significance levels, different models, and different types of traits. In our simulations we consider 10 traits (K = 10). In each simulation scenario, the p-values of TOWmuT are estimated by 1000 permutations and the type I error rates of TOWmuT are evaluated using 10,000 replicated samples. For 10,000 replicated samples, the 95% confidence intervals (CIs) for the estimated type I error rates of nominal levels 0.05 and 0.01 are (0.046, 0.054) and (0.008, 0.012), respectively. The estimated type I error rates of TOWmuT are summarized in Tables 3.1 and 3.2. From these two tables, we can see that 70 out of 72 (greater than 95%) estimated type I error rates are within the 95% CIs and the two estimated type I error rates not within the 95% CIs (0.05555 and 0.01295) are very close to the bound of the corresponding 95% CI, which indicates that TOWmuT is valid.

For power comparisons, we consider different values of heritability, different models, different types of traits, different percentages of protective variants, different values of between-factor correlation, and different values of within-factor correlation. In each of the simulation scenarios, the p-values of TOWmuT, AWRR and single-TOW are estimated using 1,000 permutations and the p-values of MANOVA, GAMuT, and MSKAT are estimated using asymptotic distributions. The powers of all of the six tests are evaluated using 1,000 replicated samples at a significance level of 0.05.

Figure 3.1 gives the power comparisons of the six tests (Single-TOW, MSKAT, AWRR, MANOVA, GAMuT, and TOWmuT) for the power as a function of the total heritability based on the six models for 10 quantitative traits. This figure shows that (1) TOWmuT is consistently the most powerful one among the six tests; (2) MANOVA is the second most powerful when genotypes impact on multiple traits (models 1-3 and 6) while AWRR is the second most powerful when genotypes impact on a single trait (models 4-5); (3) MSKAT is consistently less powerful than other multivariate tests probably because SKAT gives larger weights than that of TOW to only those variants with MAF in the range (0.01,0.035) and there are only 8% variants with MAF in the range (0.01,0.035) in Sgene which our simulations are based on; and (4) MSKAT and GAMuT have similar powers in all six models.

Figure 3.2 gives the power comparisons of the five tests (Single-TOW, AWRR, MSKAT, GAMuT, and TOWmuT) for the power as a function of the total heritability for the mixture of 5 quantitative traits and 5 qualitative traits. We only compare the powers of five tests because MANOVA has inflated type I error rate in this case. This figure shows that (1) TOWmuT is consistently the most powerful one among the five tests; (2) AWRR is second most powerful when genotypes impact on multiple traits (models 1-3 and 6) while MSKAT and GAMuT are second most powerful when genotypes impact on a single trait (models 4-5); (3) MSKAT and GAMuT have similar powers in all six models; and (4) single-TOW is consistently less powerful than other four multivariate tests because we

keep correlations between traits similar to that in Figure 3.1 such that correlations between original quantitative traits are larger than that in Figure 3.1.

We also compare the powers of the six tests for the power as a function of the within-factor correlation for models 1-5 and between-factor correlation for model 6 for 10 quantitative traits (Figure B.3.1). As shown in this figure, the power of single-TOW is robust to the between-factor correlation or the within-factor correlation since the minimum p-value-based approach is largely unaffected by the trait correlation (Wu and Pankow 2016). However, with the increasing of the between-factor correlation or within-factor correlation, the power of other five tests essentially increases. Other patterns of the power comparisons are similar to those of in Figure B.3.1.

Power comparisons of the six tests for the power as a function of the percentage of protective variants for 10 quantitative traits are given by Figure B.3.2. This figure shows that the power of all six tests are robust to the percentage of protective variants, therefore, all of these methods are robust to the directions of the genetic effects. Other patterns of the power comparisons are similar to those of in Figure 3.1.

Application to the COPDGene

Chronic obstructive pulmonary disease (COPD) is a common disease in elderly patients that causes significant morbidity and mortality [Nazir and Erbland, 2009]. The Genetic Epidemiology of COPD Study (COPDGene) [Regan et al., 2010] was designed to identify genetic factors associated with COPD. In this COPDGene study, a total of more than 10,000 subjects have been enrolled including 2/3 non-Hispanic Whites (NHW) and 1/3 African-Americans (AA). In this analysis, we only include 5,430 NHW with no missing phenotypes. Each of the 5,430 NHW has been genotyped at 630,860 SNPs. Based on the literature studies of COPD [Han et al., 2011; Chu et al., 2014; Liang et al., 2016], we selected 7 key quantitative COPD-related phenotypes, including FEV1 (% predicted FEV1), Emphysema (Emph), Emphysema Distribution (EmphDist), Gas Trapping (GasTrap), Airway Wall Area (Pi10), Exacerbation frequency (ExacerFreq), Six-minute walk distance (6MWD), and 4 covariates, including BMI, Age, Pack-Years (PackYear) and Sex.

To evaluate the performance of our proposed method on a real data set, we applied six methods (TOWmuT, MANOVA, MSKAT, GAMuT, AWRR, and single-TOW) to the COPDGene of NHW population to test the association between each of 50-SNP blocks and the 7 key quantitative COPD-related phenotypes. To identify significant 50-SNP blocks associated with the phenotypes, we used Bonferroni correction to decide the significance level. The total number of 50-SNP blocks is 12617, therefore, the Bonferroni corrected significance level is $0.05/12617 \approx 4\times10^{-6}$. Table 3.3 summarized the significant blocks identified by at least one method. There were total six significant blocks in Table 3. All of the six blocks have been previously reported to be in association with COPD or lung functions [Pillai et al., 2009; Cho et al., 2010; Figarska et al., 2014; Lutz et al., 2015]. PDSS1 and ABI1 are located between LOC107984176 and LOC105376467, which are Intergenic regions and contain the SNPs associated with pulmonary function [Imboden et al., 2012; Lutz et al., 2015]. From Table 3, we can see that TOWmuT identified four blocks; AWRR identified two blocks; MANOVA, MSKAT and GAMuT identified one block; single-TOW did not identify any blocks. From these results, we can see that TOWmuT identified the most of significant 50-SNP blocks among the six methods, which is consistent with the results of our simulation studies.

Discussion

We developed TOWmuT to perform joint analysis of multiple traits in gene-based association studies based on the following reasons: (1) multiple related traits are usually measured in genetic association studies of complex diseases; (2) there is increasing evidence showing that pleiotropy is a widespread phenomenon in complex diseases; and (3) there is a shortage of gene-based approaches for multiple traits. We used extensive simulation studies to compare the performance of TOWmuT with MANOVA, MSKAT, AWRR, GAMuT and Single-TOW. Our results showed that TOWmuT has correct type I error rates and is consistently more powerful than other five methods. Additionally, the real data analysis results demonstrated that the proposed method has great potential in gene-based association study for complex diseases with multiple phenotypes such as COPD.

Recently, it has become a major focus of investigation to identify a small number of rare causal variants that contribute to complex diseases [Capanu and Ionita-Laza, 2015]. Several methods to pinpoint the causal variants have been developed for testing the association with a single trait. These methods include backward elimination (BE) method [Ionita-Laza et al., 2014], hierarchical model method [Ionita-Laza et al., 2014], and adaptive combination of p-values method [Lin, 2016]. To extend the TOWmuT method to identify a small number of causal variants which are associated with multiple traits, we can use the BE method. In each step, we remove one variant that has the smallest contribution to the association between multiple traits and the set of variants and then we evaluate the p-value for testing association between multiple traits and the remaining variants by TOWmuT. Causal variants are the set of variants corresponding to the smallest p-value.

The computation time required for running TOWmuT depends on the sample size, the number of variants in the genomic region, the number of traits, and the number of permutations. The running time of TOWmuT with 1000 permutations on a data set with 5000 individuals, 7 traits and 10 variants in a genomic region on a laptop with 4 Intel Cores @ 3.30GHz and 4 GB memory is about 0.14s. To perform genome-wide association studies, we can first select genomic regions that show evidence of association based on a small number of permutations (e.g. 1,000), and then a large number of permutations are used to test the selected regions.

Tables and Figures

Table 3.1. The estimated type I error rates of TOWmuT for 10 quantitative traits under each model with covariates.

	Sample Size			
	Model	500	1000	2000
$\alpha = 0.05$	1	0.05365	0.0515	0.0515
	2	0.0521	0.0528	0.0504
	3	0.0513	0.0540	0.0503
	4	0.0514	0.0511	0.05
	5	0.05381	0.04825	0.05
	6	0.0482	0.0508	0.05325
$\alpha = 0.01$	1	0.01165	0.0098	0.0117
	2	0.012	0.01015	0.0102
	3	0.01175	0.01075	0.0113
	4	0.01145	0.01075	0.0118
	5	0.01141	0.01095	0.0117
	6	0.0097	0.0105	0.01185

Table 3.2. The estimated type I error rates of TOWmuT for the mixture of five quantitative traits and five qualitative traits under each model with covariates.

	Sample Size			
	Model	500	1000	2000
$\alpha = 0.05$	1	0.05365	0.05385	0.05005
	2	0.0511	0.0483	0.05115
	3	0.0508	0.05375	0.052
	4	0.0529	0.04915	0.0536
	5	0.054	0.05355	0.04825
	6	0.05555	0.0493	0.0529
$\alpha = 0.01$	1	0.0105	0.01295	0.00995
	2	0.0105	0.009	0.0097
	3	0.01145	0.0104	0.0101
	4	0.01065	0.00945	0.01165
	5	0.0118	0.0105	0.00875
	6	0.01195	0.00935	0.01105

Table 3.3. Significant blocks identified by at least one method (p-values less than 4×10-6) and the corresponding p-values in the analysis of COPDGene.

CHR	POS1	POS2	Genes	TOWmuT	MANOVA	MSKAT	GAMuT	AWRR	Single-TOW
2	178000985	178419117	NFE2L2	0.20883	2.62E-06	0.02508	0.02505	0.25796	0.15468
4	145278837	145697040	HHIP	1.00E-07	7.71E-06	0.03992	0.03984	0	0.00085
10	26908475	27150093	PDSS1, ABI1	4.00E-06	0.04050	0.01242	0.01247	1.6E-05	0.02845
15	78593362	78825917	IREB2, AGPHD1	1.00E-07	0.00191	0.70349	0.70357	5.6E-06	0.23484
15	78826180	79006442	PSMA4, CHRNA5, CHRNA3, CHRNB4	2.90E-06	0.00037	0.06255	0.06252	0	0.37643
15	79006582	79267817	ADAMTS7	9.01E-05	4.78E-05	2.25E-06	6.42E-07	0.04849	0.01953

Figure 3.1. Power comparisons of the six tests (Single-TOW, MSKAT, AWRR, MANOVA, GAMuT and TOWmuT) for the power as a function of total heritability for 10 quantitative traits with covariates. The sample size is 1000. The between-factor correlation is 0.3 and the within-factor correlation is 0.7. The percentage of the causal variants is 0.2. All causal variants are risk variants.

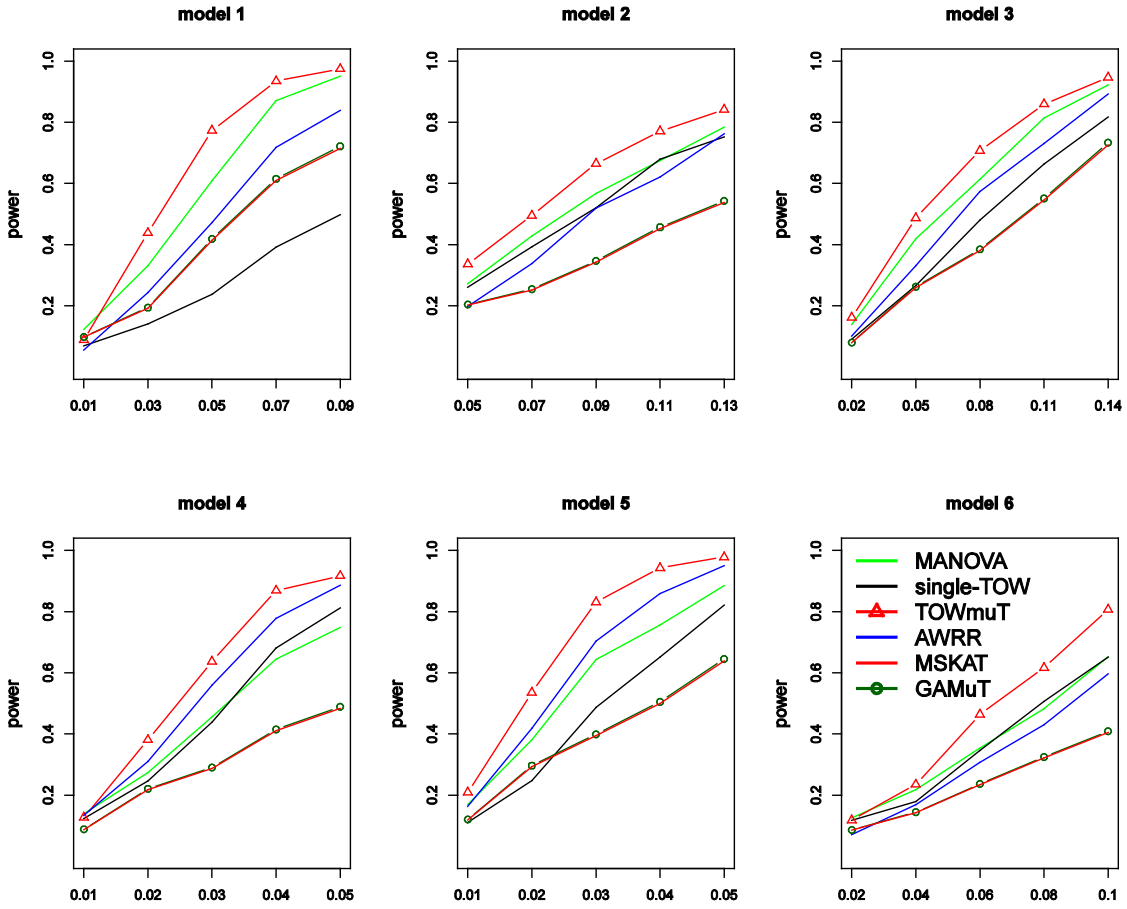
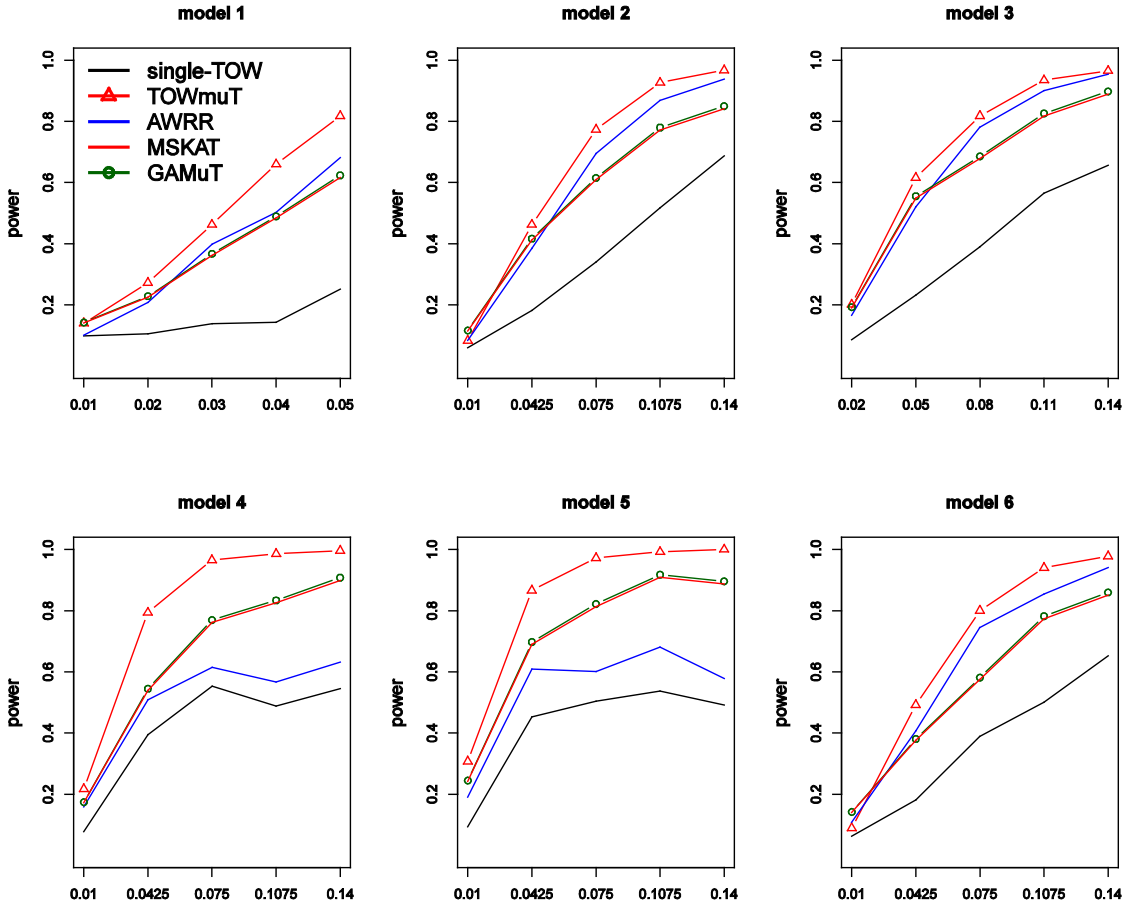


Figure 3.2. Power comparisons of the five tests (Single-TOW, AWRR, GAMuT, MSKAT and TOWmuT) for the power as a function of heritability for the mixture of half quantitative traits and half qualitative traits with covariates. The sample size is 1000. Covariance matrix of 10 traits is similar to that of 10 quantitative traits with between-factor correlation being 0.3 and the within-factor correlation being 0.7. The percentage of the causal variants is 0.2. All causal variants are risk variants.



3 Reference List

Andres AM, Clark AG, Shimmin L, Boerwinkle E, Sing CF, Hixson JE. 2007. Understanding the accuracy of statistical haplotype inference with sequence data of known phase. Genet Epidemiol 31(7):659-71.

Aschard H, Vilhjalmsson BJ, Greliche N, Morange PE, Tregouet DA, Kraft P. 2014. Maximizing the power of principal-component analysis of correlated phenotypes in genome-wide association studies. Am J Hum Genet 94(5):662-76.

Bauchet M, McEvoy B, Pearson LN, Quillen EE, Sarkisian T, Hovhannesyan K, Deka R, Bradley DG, Shriver MD. 2007. Measuring European population stratification with microarray genotype data. Am J Hum Genet 80(5):948-56.

Bodmer W, Bonilla C. 2008. Common and rare variants in multifactorial susceptibility to common diseases. Nat Genet 40(6):695-701.

Broadaway KA, Cutler DJ, Duncan R, Moore JL, Ware EB, Jhun MA, Bielak LF, Zhao W, Smith JA, Peyser PA and others. 2016. A Statistical Approach for Testing Cross-Phenotype Effects of Rare Variants. Am J Hum Genet 98(3):525-40.

Capanu M, Ionita-Laza I. 2015. Integrative analysis of functional genomic annotations and sequencing data to identify rare causal variants via hierarchical modeling. Front Genet 6:17.

Casale FP, Rakitsch B, Lippert C, Stegle O. 2015. Efficient set tests for the genetic analysis of correlated traits. Nat Methods 12(8):755-8.

Chen HS, Zhu X, Zhao H, Zhang S. 2003. Qualitative semi-parametric test for genetic associations in case-control designs under structured populations. Ann Hum Genet 67(Pt 3):250-64.

Cho MH, Boutaoui N, Klanderman BJ, Sylvia JS, Ziniti JP, Hersh CP, DeMeo DL, Hunninghake GM, Litonjua AA, Sparrow D and others. 2010. Variants in FAM13A are associated with chronic obstructive pulmonary disease. Nat Genet 42(3):200-2.

Chu JH, Hersh CP, Castaldi PJ, Cho MH, Raby BA, Laird N, Bowler R, Rennard S, Loscalzo J, Quackenbush J and others. 2014. Analyzing networks of phenotypes in complex diseases: methodology and applications in COPD. BMC Syst Biol 8:78.

Derkach A, Lawless JF, Sun L. 2013. Robust and powerful tests for rare variants using Fisher's method to combine evidence of association from two or more complementary tests. Genet Epidemiol 37(1):110-21.

Devlin B, Roeder K. 1999. Genomic control for association studies. Biometrics 55(4):997-1004.

Devlin B, Roeder K, Wasserman L. 2001. Genomic control, a new approach to genetic-based association studies. Theor Popul Biol 60(3):155-66.

Falconer DS, Mackay TFC. 1996. Introduction to quantitative genetics. Essex, England: Longman.

Feng T, Zhang S, Sha Q. 2007. A method dealing with a large number of correlated traits in a linkage genome scan. BMC Proc 1 Suppl 1:S84.

Ferreira MA, Purcell SM. 2009. A multivariate test of association. Bioinformatics 25(1):132-3.

Figarska SM, Vonk JM, Boezen HM. 2014. NFE2L2 polymorphisms, mortality, and metabolism in the general population. Physiol Genomics 46(12):411-7.

Ge Y, Dudoit S, Speed TP. 2003. Resampling-based multiple testing for microarray data analysis. Test 12(1):1-77.

Greco B, Hainline A, Arbet J, Grinde K, Benitez A, Tintle N. 2015. A general approach for combining diverse rare variant association tests provides improved robustness across a wider range of genetic architectures. Eur J Hum Genet.

Gur RE, Nimgaonkar VL, Almasy L, Calkins ME, Ragland JD, Pogue-Geile MF, Kanes S, Blangero J, Gur RC. 2007. Neurocognitive endophenotypes in a multiplex multigenerational family study of schizophrenia. Am J Psychiatry 164(5):813-9.

Han MK, Kazerooni EA, Lynch DA, Liu LX, Murray S, Curtis JL, Criner GJ, Kim V, Bowler RP, Hanania NA and others. 2011. Chronic obstructive pulmonary disease exacerbations in the COPDGene study: associated radiologic phenotypes. Radiology 261(1):274-82.

Imboden M, Bouzigon E, Curjuric I, Ramasamy A, Kumar A, Hancock DB, Wilk JB, Vonk JM, Thun GA, Siroux V. 2012. Genome-wide association study of lung function decline in adults with and without asthma. Journal of allergy and clinical immunology 129(5):1218-1228.

Ionita-Laza I, Capanu M, De Rubeis S, McCallum K, Buxbaum JD. 2014. Identification of rare causal variants in sequence-based studies: methods and applications to VPS13B, a gene involved in Cohen syndrome and autism. PLoS Genet 10(12):e1004729.

Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB, Sabatti C, Eskin E. 2010. Variance component model to account for sample structure in genome-wide association studies. Nat Genet 42(4):348-54.

Kim J, Bai Y, Pan W. 2015a. An Adaptive Association Test for Multiple Phenotypes with GWAS Summary Statistics. Genet Epidemiol.

Kim J, Bai Y, Pan W. 2015b. An Adaptive Association Test for Multiple Phenotypes with GWAS Summary Statistics. Genetic Epidemiology 39(8):651-663.

Klei L, Luca D, Devlin B, Roeder K. 2008. Pleiotropy and principal components of heritability combine to increase power for association analysis. Genet Epidemiol 32(1):9-19.

Knowler WC, Williams RC, Pettitt DJ, Steinberg AG. 1988. Gm3;5,13,14 and type 2 diabetes mellitus: an association in American Indians with genetic admixture. Am J Hum Genet 43(4):520-6.

Korte A, Vilhjalmsson BJ, Segura V, Platt A, Long Q, Nordborg M. 2012. A mixed-model approach for genome-wide association studies of correlated traits in structured populations. Nat Genet 44(9):1066-71.

Lander ES, Schork NJ. 1994. Genetic dissection of complex traits. Science 265(5181):2037-48.

Lange C, van Steen K, Andrew T, Lyon H, DeMeo DL, Raby B, Murphy A, Silverman EK, MacGregor A, Weiss ST and others. 2004. A family-based association test for repeatedly measured quantitative traits adjusting for unknown environmental and/or polygenic effects. Stat Appl Genet Mol Biol 3:Article17.

Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, Team NGESP-ELP, Christiani DC, Wurfel MM, Lin X. 2012. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. Am J Hum Genet 91(2):224-37.

Lee S, Teslovich TM, Boehnke M, Lin X. 2013. General framework for meta-analysis of rare variants in sequencing association studies. Am J Hum Genet 93(1):42-53.

Li B, Leal SM. 2008. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. Am J Hum Genet 83(3):311-21.

Liang X, Wang Z, Sha Q, Zhang S. 2016. An Adaptive Fisher's Combination Method for Joint Analysis of Multiple Phenotypes in Association Studies. Sci Rep 6:34323.

Lin WY. 2016. Beyond Rare-Variant Association Testing: Pinpointing Rare Causal Variants in Case-Control Sequencing Study. Sci Rep 6:21824.

Lutz SM, Cho MH, Young K, Hersh CP, Castaldi PJ, McDonald ML, Regan E, Mattheisen M, DeMeo DL, Parker M and others. 2015. A genome-wide association study identifies risk loci for spirometric measures among smokers of European and African ancestry. BMC Genet 16:138.

Madsen BE, Browning SR. 2009. A groupwise association test for rare mutations using a weighted sum statistic. PLoS Genet 5(2):e1000384.

Metzker ML. 2010. Sequencing technologies - the next generation. Nat Rev Genet 11(1):31-46.

Morgenthaler S, Thilly WG. 2007. A strategy to discover genes that carry multi-allelic or mono-allelic risk for common diseases: a cohort allelic sums test (CAST). Mutat Res 615(1-2):28-56.

Nazir SA, Erbland ML. 2009. Chronic obstructive pulmonary disease: an update on diagnosis and management issues in older adults. Drugs Aging 26(10):813-31.

Neale BM, Rivas MA, Voight BF, Altshuler D, Devlin B, Orho-Melander M, Kathiresan S, Purcell SM, Roeder K, Daly MJ. 2011. Testing for an unusual distribution of rare variants. PLoS Genet 7(3):e1001322.

Newton-Cheh C, Johnson T, Gateva V, Tobin MD, Bochud M, Coin L, Najjar SS, Zhao JH, Heath SC, Eyheramendy S and others. 2009. Genome-wide association study identifies eight loci associated with blood pressure. Nat Genet 41(6):666-76.

O'Brien PC. 1984. Procedures for comparing samples with multiple endpoints. Biometrics 40(4):1079-87.

O'Reilly PF, Hoggart CJ, Pomyen Y, Calboli FC, Elliott P, Jarvelin MR, Coin LJ. 2012. MultiPhen: joint model of multiple phenotypes can increase discovery in GWAS. PLoS One 7(5):e34861.

Ott J, Rabinowitz D. 1999. A principal-components approach based on heritability for combining phenotype information. Hum Hered 49(2):106-11.

Pillai SG, Ge D, Zhu G, Kong X, Shianna KV, Need AC, Feng S, Hersh CP, Bakke P, Gulsvik A and others. 2009. A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. PLoS Genet 5(3):e1000421.

Price AL, Kryukov GV, de Bakker PI, Purcell SM, Staples J, Wei LJ, Sunyaev SR. 2010. Pooled association tests for rare variants in exon-resequencing studies. Am J Hum Genet 86(6):832-8.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38(8):904-9.

Pritchard JK. 2001. Are rare variants responsible for susceptibility to complex diseases? Am J Hum Genet 69(1):124-37.

Pritchard JK, Cox NJ. 2002. The allelic architecture of human disease genes: common disease-common variant...or not? Hum Mol Genet 11(20):2417-23.

Regan EA, Hokanson JE, Murphy JR, Make B, Lynch DA, Beaty TH, Curran-Everett D, Silverman EK, Crapo JD. 2010. Genetic epidemiology of COPD (COPDGene) study design. COPD 7(1):32-43.

Reich DE, Goldstein DB. 2001. Detecting association in a case-control study while correcting for population stratification. Genet Epidemiol 20(1):4-16.

Rifai N, Ridker PM. 2002. Inflammatory markers and coronary heart disease. Curr Opin Lipidol 13(4):383-9.

Sha Q, Wang X, Wang X, Zhang S. 2012. Detecting association of rare and common variants by testing an optimally weighted combination of variants. Genet Epidemiol 36(6):561-71.

Sha Q, Zhang S. 2014. A rare variant association test based on combinations of single-variant tests. Genet Epidemiol 38(6):494-501.

Sivakumaran S, Agakov F, Theodoratou E, Prendergast JG, Zgaga L, Manolio T, Rudan I, McKeigue P, Wilson JF, Campbell H. 2011. Abundant pleiotropy in human complex diseases and traits. Am J Hum Genet 89(5):607-18.

Solovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW. 2013. Pleiotropy in complex traits: challenges and strategies. Nat Rev Genet 14(7):483-95.

Stephens M. 2013. A unified framework for association analysis with multiple related phenotypes. PLoS One 8(7):e65245.

Stratton MR, Rahman N. 2008. The emerging landscape of breast cancer susceptibility. Nat Genet 40(1):17-22.

Sun J, Oualkacha K, Forgetta V, Zheng HF, Brent Richards J, Ciampi A, Greenwood CM, Consortium UK. 2016. A method for analyzing multiple continuous phenotypes in rare variant association studies allowing for flexible correlations in variant effects. Eur J Hum Genet.

Tang CS, Ferreira MA. 2012. A gene-based test of association using canonical correlation analysis. Bioinformatics 28(6):845-50.

Teer JK, Mullikin JC. 2010. Exome sequencing: the sweet spot before whole genomes. Hum Mol Genet 19(R2):R145-51.

The UK10K Consortium, Walter K, Min JL, Huang J, Crooks L, Memari Y, McCarthy S, Perry JR, Xu C, Futema M and others. 2015. The UK10K project identifies rare variants in health and disease. Nature 526(7571):82-90.

van der Sluis S, Posthuma D, Dolan CV. 2013. TATES: efficient multivariate genotype-phenotype analysis for genome-wide association studies. PLoS Genet 9(1):e1003235.

Walsh T, King MC. 2007. Ten genes for inherited breast cancer. Cancer Cell 11(2):103-5.

Wang Y, Liu A, Mills JL, Boehnke M, Wilson AF, Bailey-Wilson JE, Xiong M, Wu CO, Fan R. 2015. Pleiotropy analysis of quantitative traits at gene level by multivariate functional linear models. Genet Epidemiol 39(4):259-75.

Wang Z, Sha Q, Zhang S. 2016a. Joint Analysis of Multiple Traits Using "Optimal" Maximum Heritability Test. PLoS One 11(3):e0150975.

Wang Z, Wang X, Sha Q, Zhang S. 2016b. Joint Analysis of Multiple Traits in Rare Variant Association Studies. Ann Hum Genet 80(3):162-71.

Wu B, Pankow JS. 2016. Sequence Kernel Association Test of Multiple Continuous Phenotypes. Genet Epidemiol 40(2):91-100.

Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. 2011. Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet 89(1):82-93.

Yan T, Li Q, Li Y, Li Z, Zheng G. 2013. Genetic association with multiple traits in the presence of population stratification. Genet Epidemiol 37(6):571-80.

Yang Q, Wang Y. 2012. Methods for analyzing multivariate phenotypes in genetic association studies. J Probab Stat 2012:652569.

Yang Q, Wu H, Guo CY, Fox CS. 2010. Analyze multivariate phenotypes in genetic association studies by combining univariate association tests. Genet Epidemiol 34(5):444-54.

Yang X, Wang S, Zhang S, Sha Q. 2017. Detecting association of rare and common variants based on cross-validation prediction error. Genet Epidemiol 41(3):233-243.

Yudkin JS, Kumari M, Humphries SE, Mohamed-Ali V. 2000. Inflammation, obesity, stress and coronary heart disease: is interleukin-6 the link? Atherosclerosis 148(2):209-14.

Zabaneh D, Balding DJ. 2010. A genome-wide association study of the metabolic syndrome in Indian Asian men. PLoS One 5(8):e11961.

Zawistowski M, Gopalakrishnan S, Ding J, Li Y, Grimm S, Zollner S. 2010. Extending rare-variant testing strategies: analysis of noncoding sequence and imputed genotypes. Am J Hum Genet 87(5):604-17.

Zeger SL, Liang KY. 1986. Longitudinal data analysis for discrete and continuous outcomes. Biometrics 42(1):121-30.

Zhang S, Zhu X, Zhao H. 2003. On a semiparametric test to detect associations between quantitative traits and candidate genes using unrelated individuals. Genet Epidemiol 24(1):44-56.

Zhang Y, Xu Z, Shen X, Pan W, Alzheimer's Disease Neuroimaging I. 2014. Testing for association with multiple traits in generalized estimation equations, with application to neuroimaging data. Neuroimage 96:309-25.

Zhang Z, Ersoz E, Lai CQ, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu J, Arnett DK, Ordovas JM and others. 2010. Mixed linear model approach adapted for genome-wide association studies. Nat Genet 42(4):355-60.

Zhou JJ, Cho MH, Lange C, Lutz S, Silverman EK, Laird NM. 2015. Integrating multiple correlated phenotypes for genetic association analysis by maximizing heritability. Hum Hered 79(2):93-104.

Zhou X, Stephens M. 2014. Efficient multivariate linear mixed model algorithms for genome-wide association studies. Nat Methods 11(4):407-9.

Zhu H, Zhang S, Sha Q. 2015a. Power Comparisons of Methods for Joint Association Analysis of Multiple Phenotypes. Hum Hered 80(3):144-52.

Zhu X, Feng T, Tayo BO, Liang J, Young JH, Franceschini N, Smith JA, Yanek LR, Sun YV, Edwards TL and others. 2015b. Meta-analysis of correlated traits via summary statistics from GWASs with an application in hypertension. Am J Hum Genet 96(1):21-36.

Zhu X, Zhang S, Zhao H, Cooper RS. 2002. Association mapping, using a mixture model for complex traits. Genet Epidemiol 23(2):181-96.

Appendix A: Supplementary Tables

Table A.1.1. The estimated type I error rates of four methods for quantitative traits and variance model 2. We use 10,000 replicates. This set of simulations is based on Sgene1.

		Sample size		
		500	1000	2000
$\alpha = 0.05$	CCA	0.049	0.05165	0.05505
	Single-TOW	0.05055	0.05325	0.05345
	WSRR	0.0472	0.0488	0.0505
	AWRR	0.0534	0.054	0.0502
$\alpha = 0.01$	CCA	0.0099	0.0116	0.0115
	Single-TOW	0.01125	0.0118	0.012
	WSRR	0.0086	0.0084	0.0103
	AWRR	0.0107	0.0114	0.0104

Table A.1.2. The estimated type I error rates of four methods for qualitative traits and variance model 2. We use 10,000 replicates. This set of simulations is based on Sgene1.

	Sample size			
		500	1000	2000
$\alpha = 0.05$	CCA	0.0473	0.05225	0.0539
	Single-TOW	0.0521	0.0559	0.0508
	WSRR	0.0485	0.0493	0.0522
	AWRR	0.0469	0.0493	0.05235
$\alpha = 0.01$	CCA	0.00975	0.0106	0.012
	Single-TOW	0.00975	0.01045	0.00955
	WSRR	0.0093	0.0109	0.0106
	AWRR	0.01115	0.0104	0.012

Appendix B: Supplementary Figures

Figure B.1.1. Power comparisons of the four tests (WSRR, AWRR, CCA and Single-TOW) for the power as a function of heritability for quantitative traits and variance model 2. The sample size is 1000 and $\rho = 0.5$. The percentage of the causal variants is 0.1. All causal variants are risk variants. The total number of traits is 10. This set of simulations is based on Sgene1.

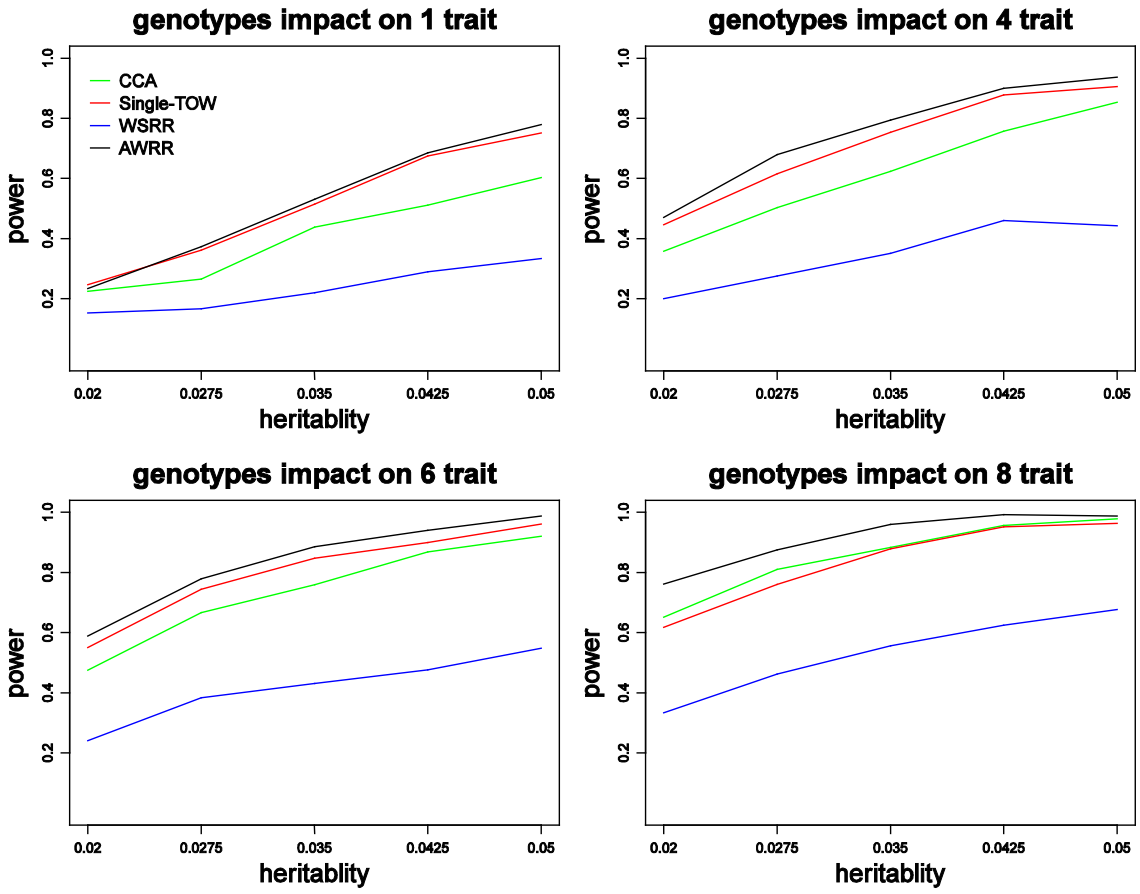


Figure B.1.2. Power comparisons of the four tests (WSRR, AWRR, CCA and Single-TOW) for the power as a function of percentage of protective variants for quantitative traits and variance model 2. The sample size is 1000, the percentage of causal variants is 0.2, the total heritability is 0.03, and $\rho = 0.5$. The total number of traits is 10. This set of simulations is based on Sgene1.

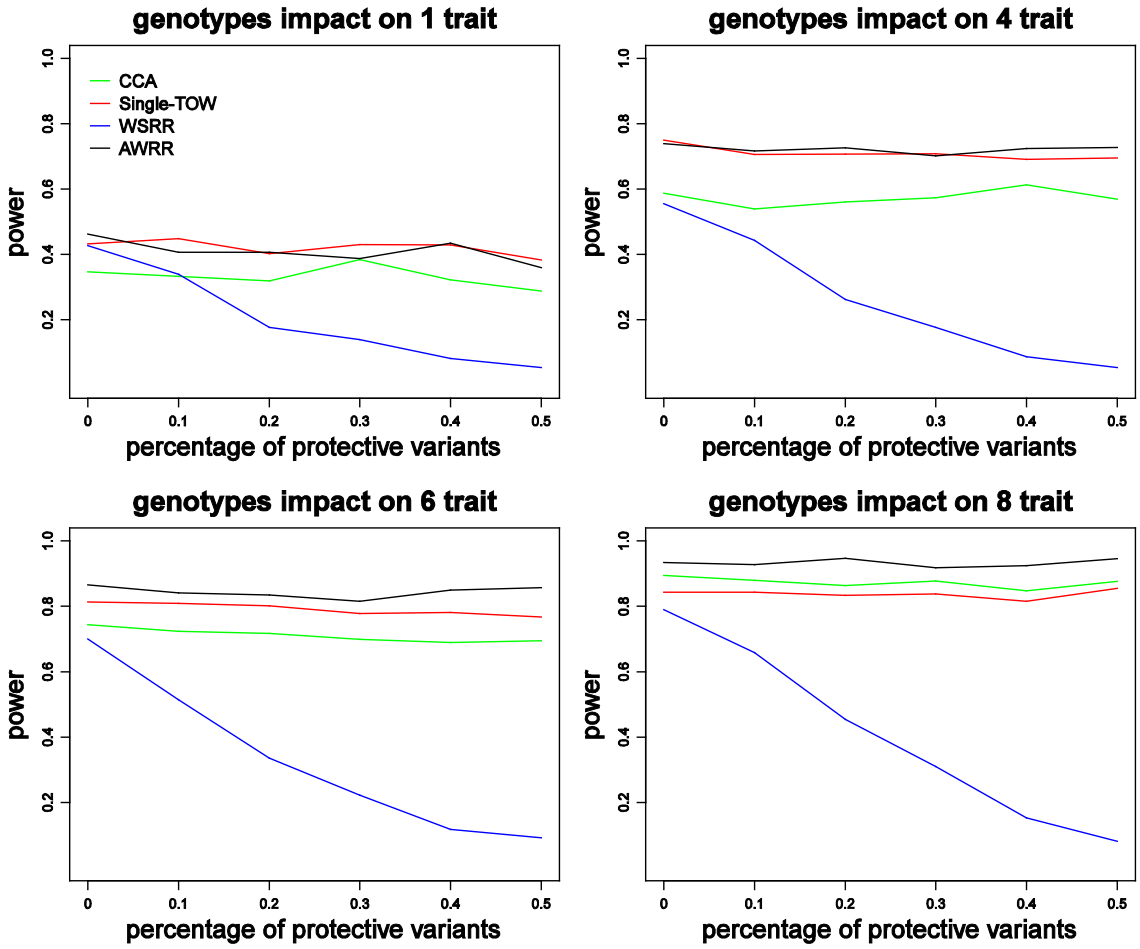


Figure B.1.3. Power comparisons of the four tests (WSRR, AWRR, CCA and Single-TOW) for the power as a function of the percentage of causal variants for quantitative traits and variance model 2. The sample size is 1000, $\rho = 0.5$, and the total heritability is 0.03. All causal variants are risk variants. The total number of traits is 10. This set of simulations is based on Sgene1.

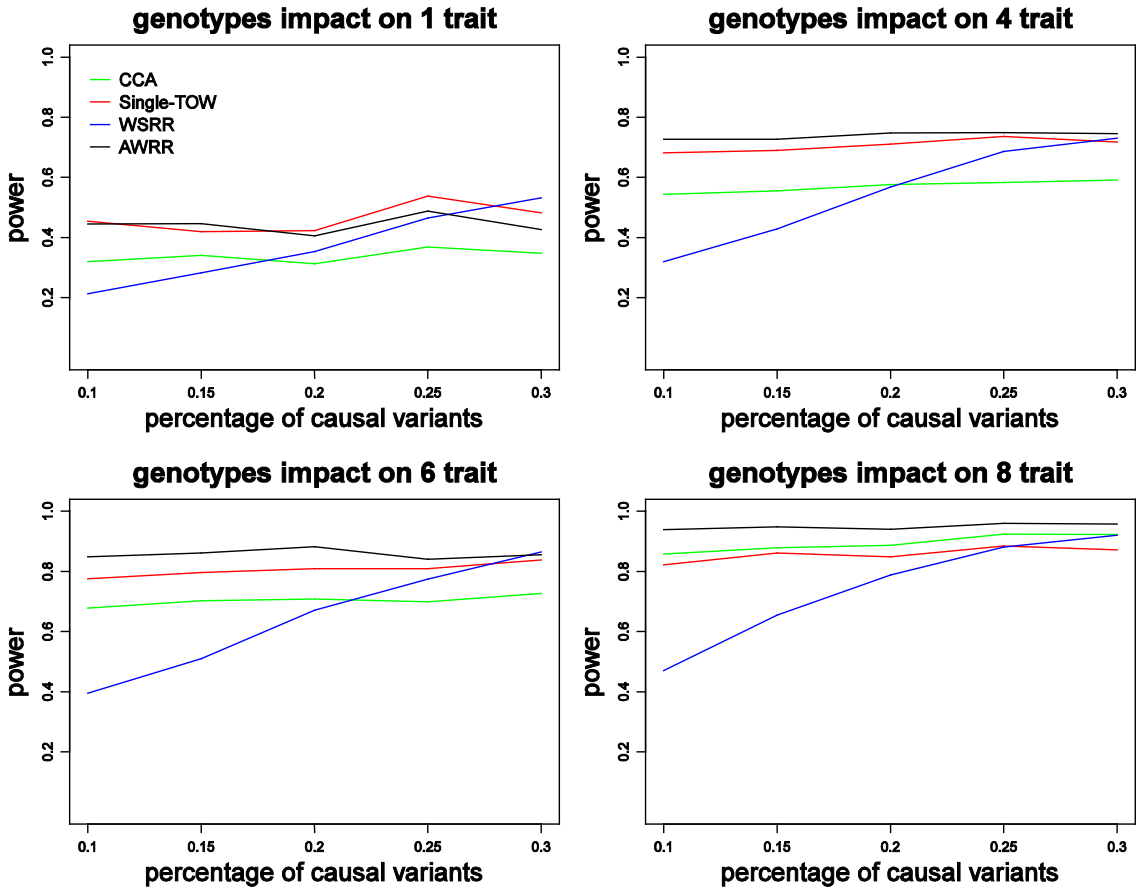


Figure B.1.4. Power comparisons of the four tests (WSRR, AWRR, CCA and Single-TOW) for the power as a function of heritability for qualitative traits and variance model 1. The sample size is 1000 and $\rho = 0.5$. The percentage of the causal variants is 0.1. All causal variants are risk variants. The total number of traits is 10. This set of simulations is based on Sgene1.

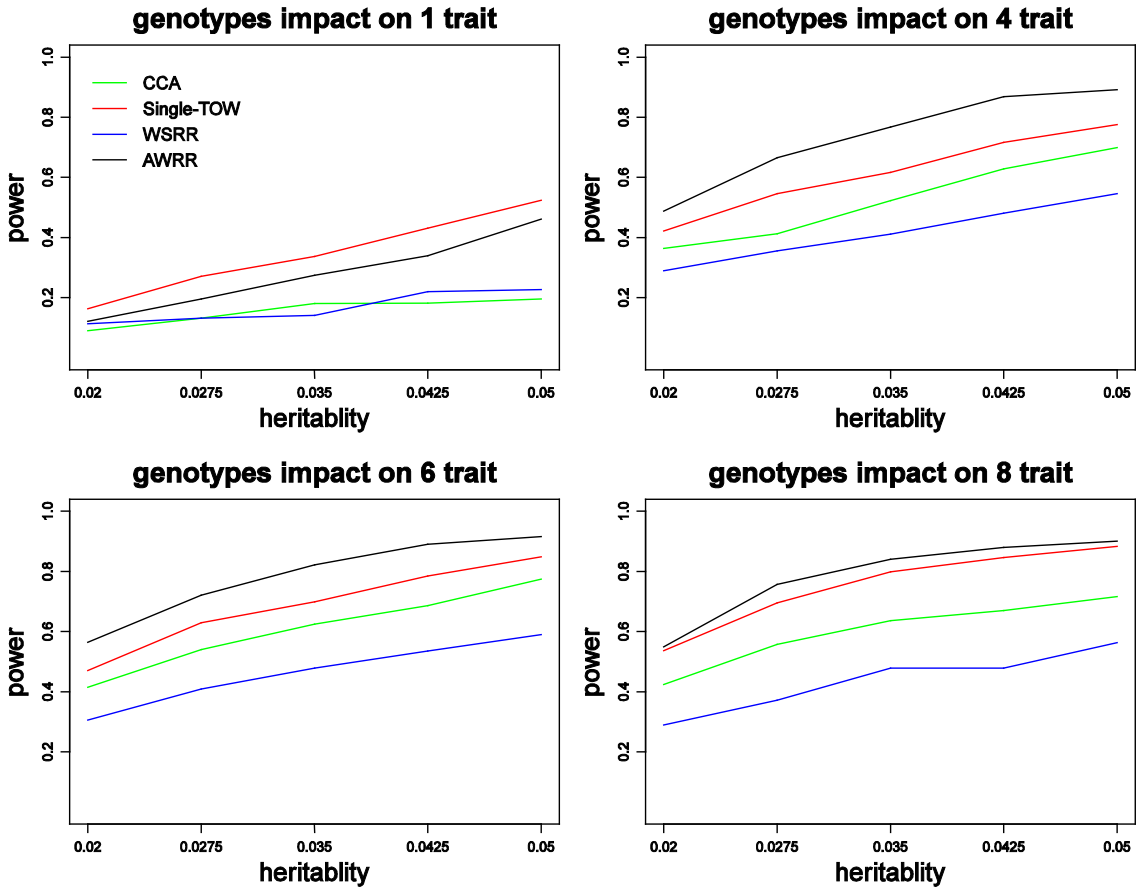


Figure B.1.5. Power comparisons of the four tests (WSRR, AWRR, CCA and Single-TOW) for the power as a function of percentage of protective variants for qualitative traits and variance model 1. The sample size is 1000, the percentage of causal variants is 0.2, the total heritability is 0.03, and $\rho = 0.5$. The total number of traits is 10. This set of simulations is based on Sgene1.

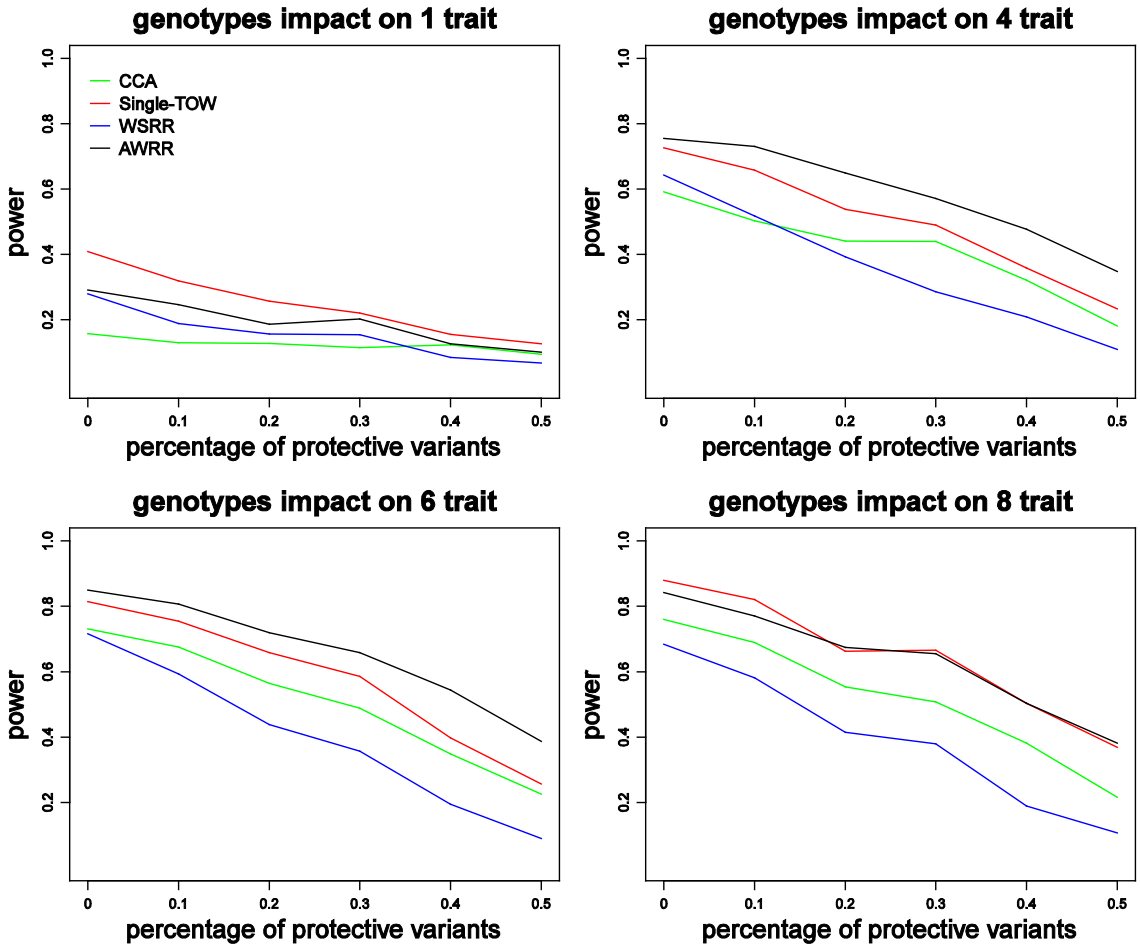


Figure B.1.6. Power comparisons of the four tests (WSRR, AWRR, CCA and Single-TOW) for the power as a function of the percentage of causal variants for qualitative traits and variance model 1. The sample size is 1000, $\rho = 0.5$, and the total heritability is 0.03. All causal variants are risk variants. The total number of traits is 10. This set of simulations is based on Sgene1.

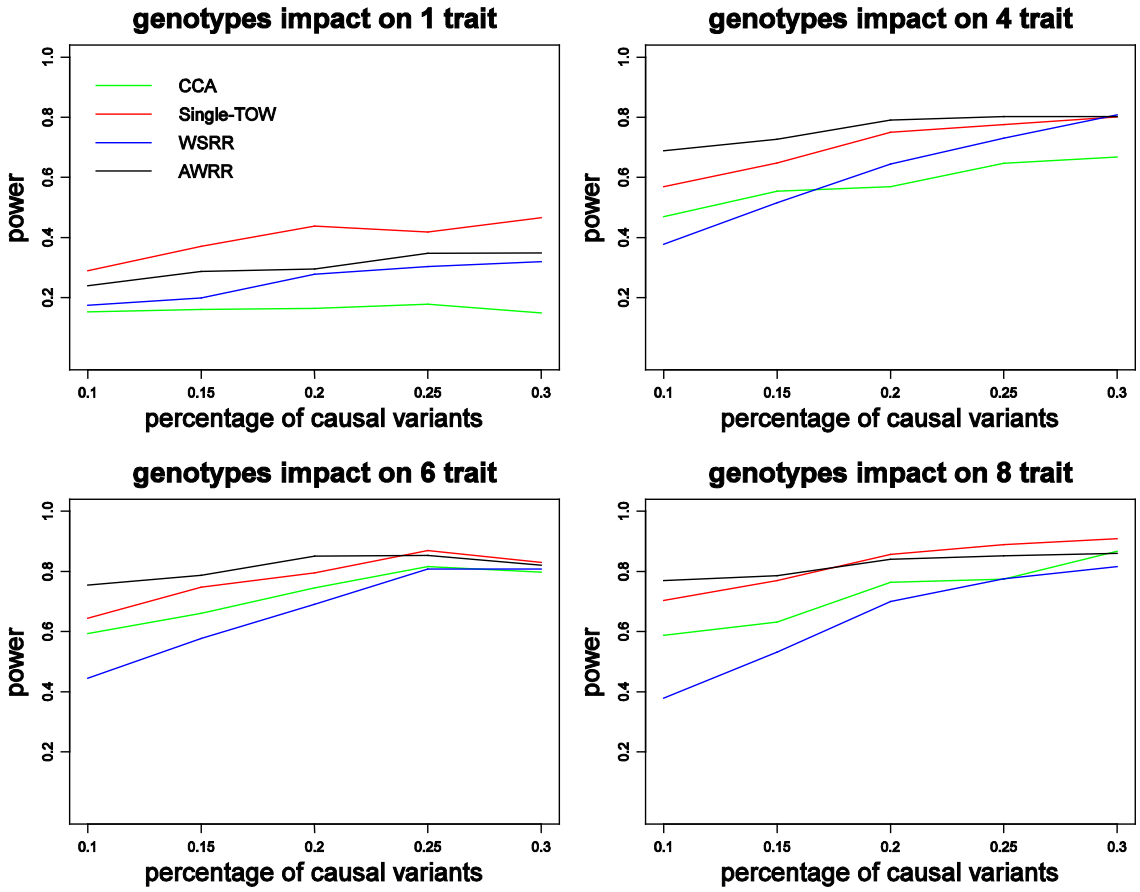


Figure B.1.7. Power comparisons of the four tests (WSRR, AWRR, CCA and Single-TOW) for the power as a function of heritability for qualitative traits and variance model 2. The sample size is 1000 and . The percentage of the causal variants is 0.1. All causal variants are risk variants. The total number of traits is 10. This set of simulations is based on Sgene1.

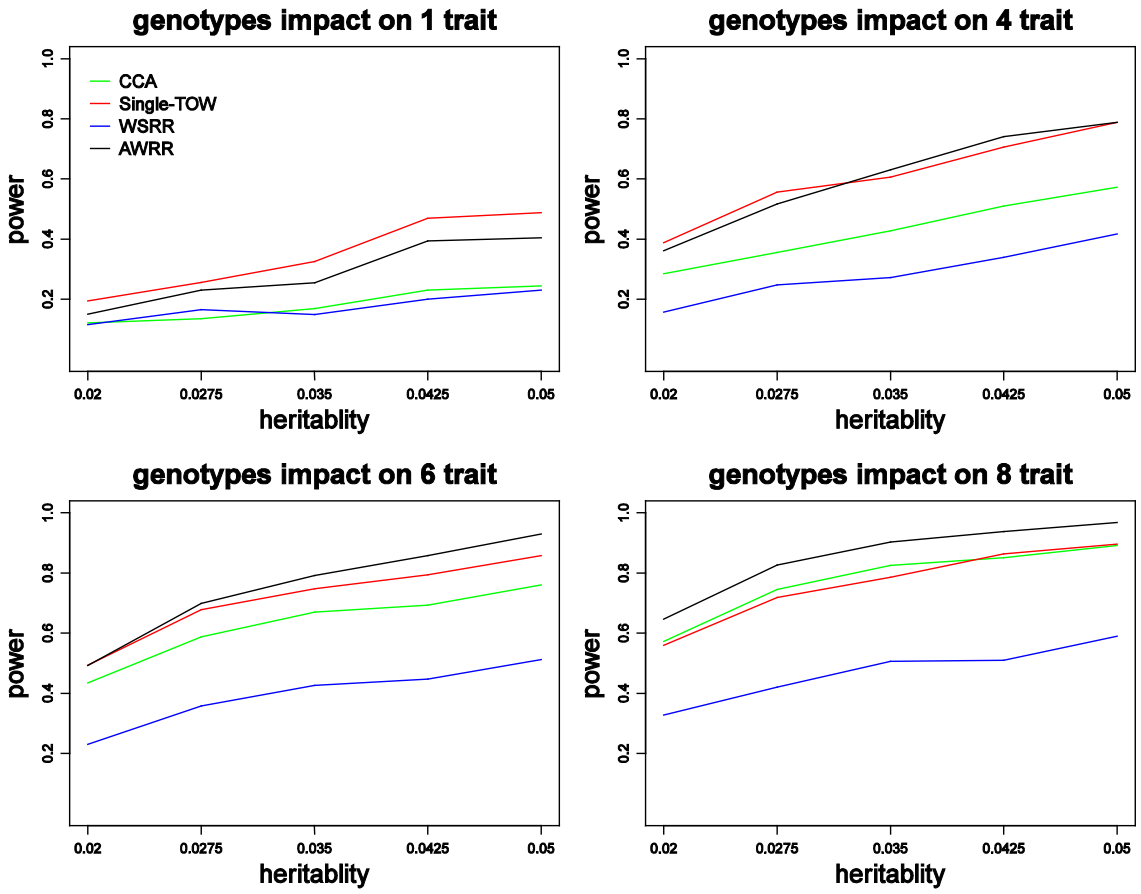


Figure B.1.8. Power comparisons of the four tests (WSRR, AWRR, CCA and Single-TOW) for the power as a function of percentage of protective variants for qualitative traits and variance model 2. The sample size is 1000, the percentage of causal variants is 0.2, the total heritability is 0.03, and $\rho = 0.5$. The total number of traits is 10. This set of simulations is based on Sgene1.

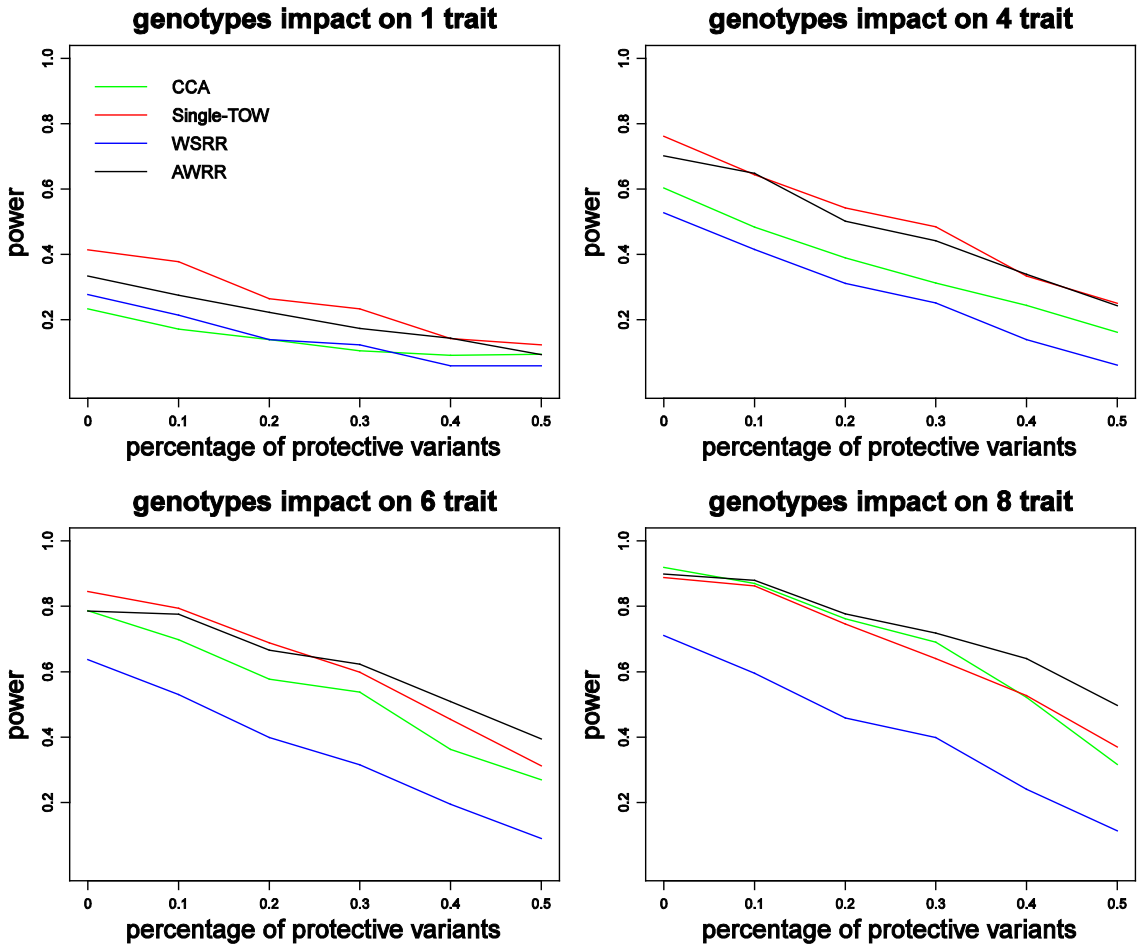


Figure B.1.9. Power comparisons of the four tests (WSRR, AWRR, CCA and Single-TOW) for the power as a function of the percentage of causal variants for qualitative traits and variance model 2. The sample size is 1000, $\rho = 0.5$, and the total heritability is 0.02. All causal variants are risk variants. The total number of traits is 10. This set of simulations is based on Sgene1.

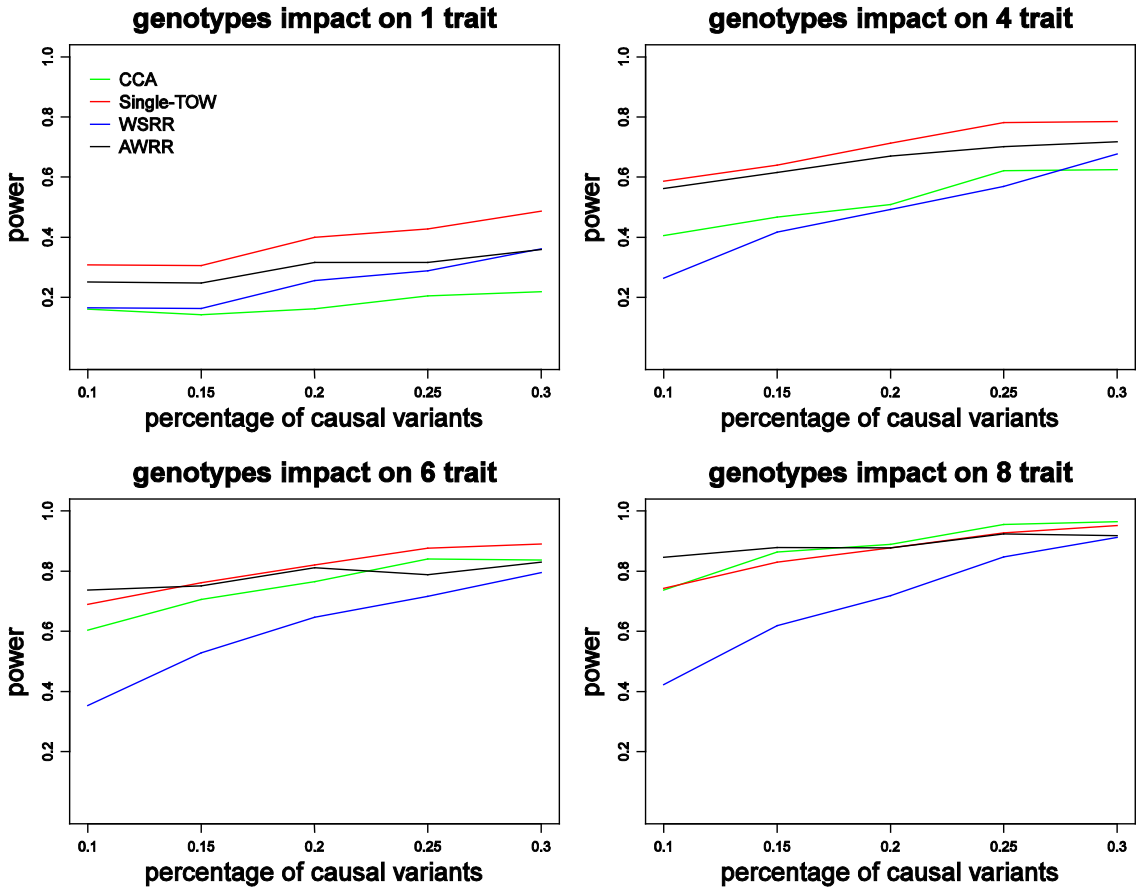


Figure B.1.10. Power comparisons of the four tests (WSRR, AWRR, CCA and Single-TOW) for the power as a function of heritability for quantitative traits under variance model 1. The sample size is 1000 and $\rho = 0.5$. The percentage of the causal variants is 0.1. All causal variants are risk variants. This set of simulations is based on Sgene2

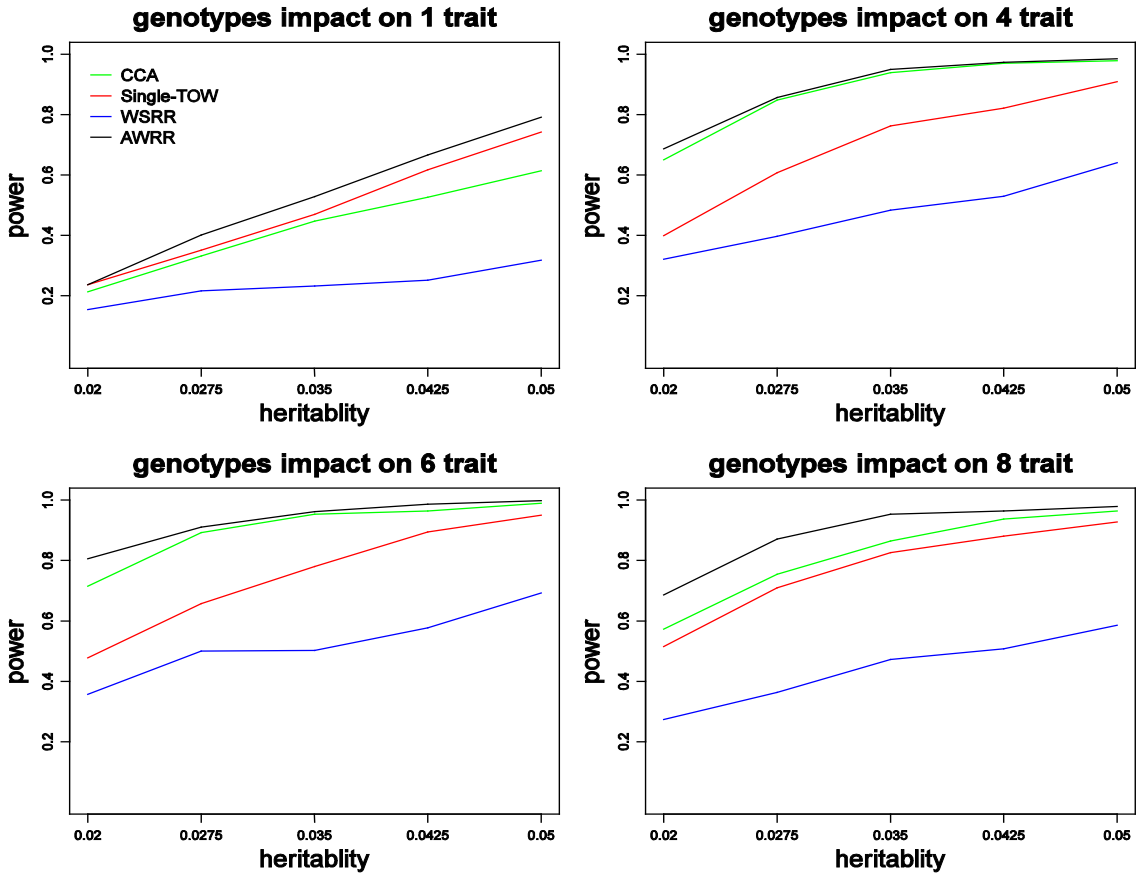


Figure B.1.11. Power comparisons of the four tests (WSRR, AWRR, CCA and Single-TOW) for the power as a function of percentage of protective variants for quantitative traits under variance model 1. The sample size is 1000, the percentage of causal variants is 0.2, the total heritability is 0.03, and $\rho = 0.5$. This set of simulations is based on Sgene2.

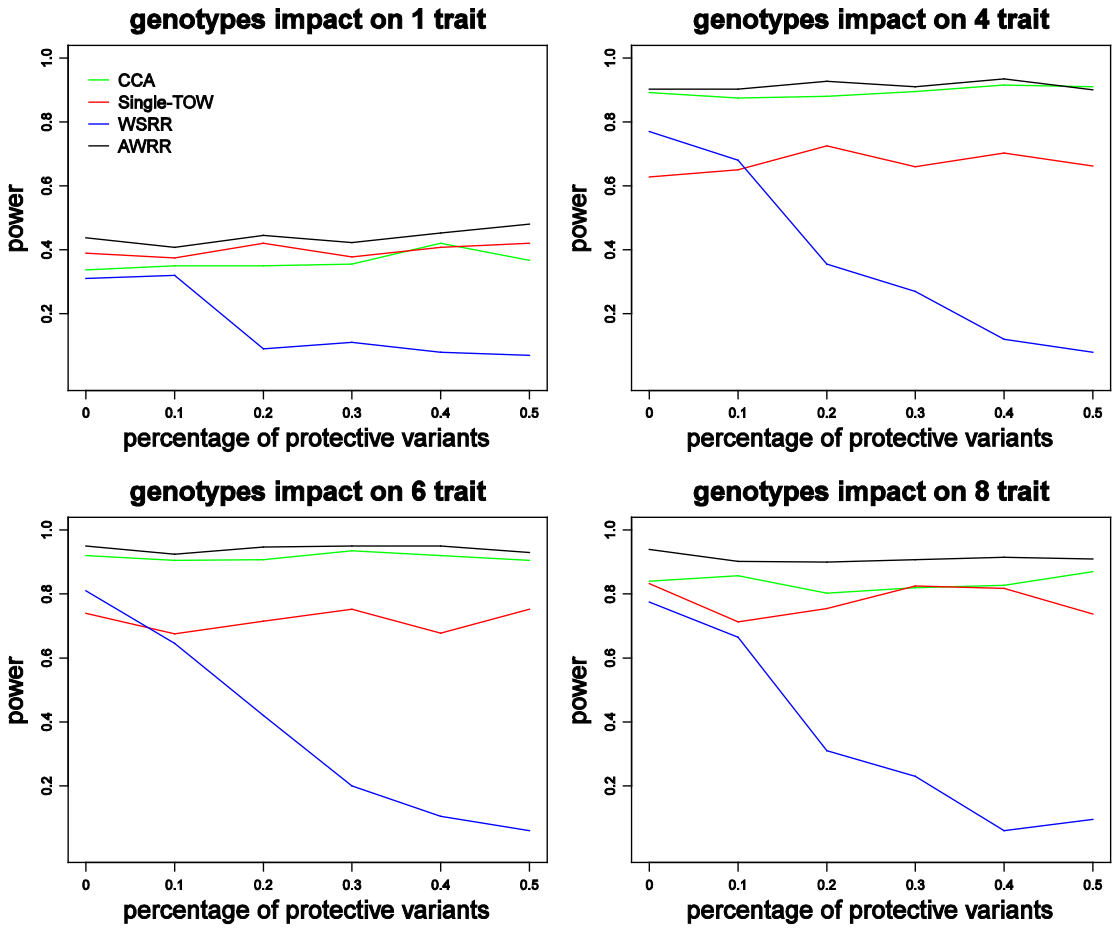


Figure B.1.12. Power comparisons of the four tests (WSRR, AWRR, CCA and Single-TOW) for the power as a function of the percentage of causal variants for quantitative traits under variance model 1. The sample size is 1000 and $\rho = 0.5$, and the total heritability is 0.03. All causal variants are risk variants. The total number of traits is 10. This set of simulations is based on Sgene2.

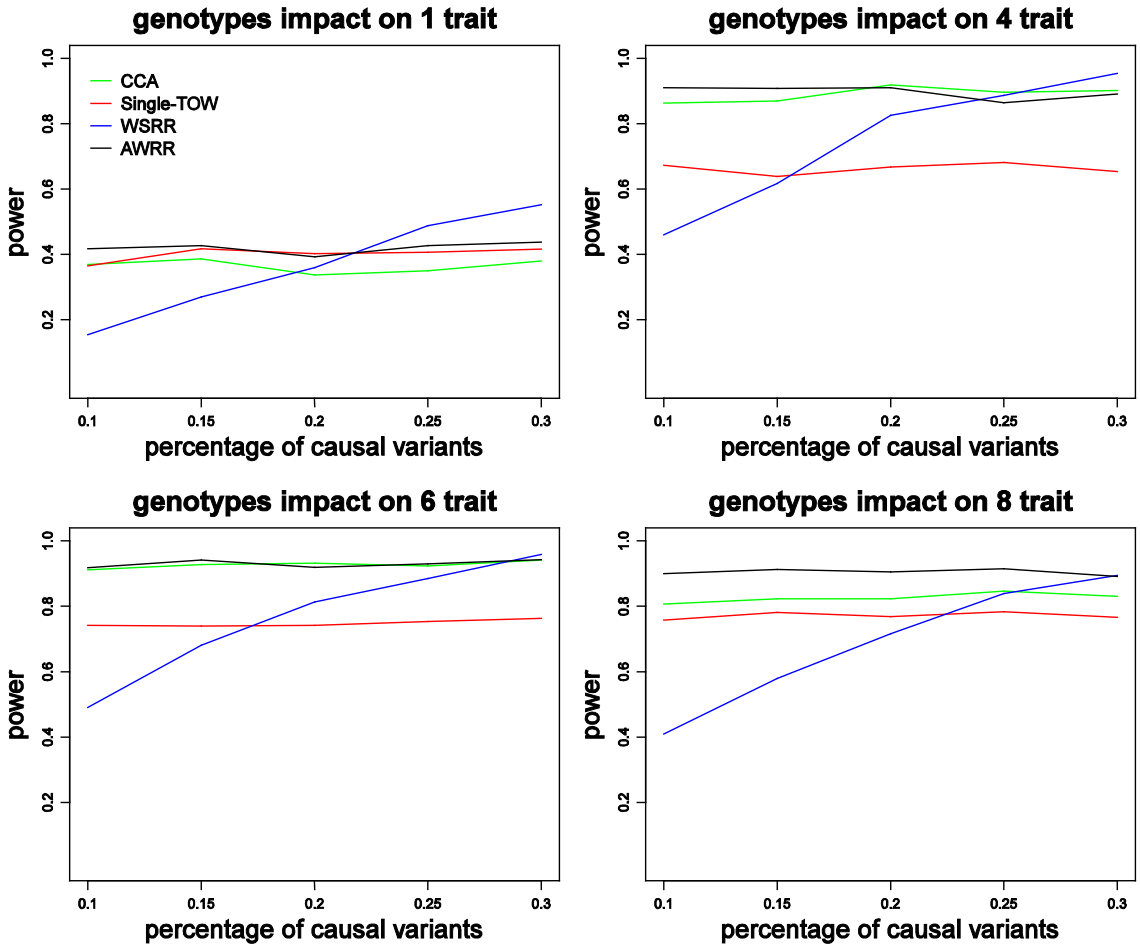


Figure B.1.13. Power comparisons of the four tests (WSRR, AWRR, CCA and Single-TOW) for the power as a function of LP for quantitative traits under variance model 1. LP represents the number of traits positively correlated with risk variants. The sample size is 1000 and the total heritability is 0.015. All causal variants are risk variants. The total number of traits is 10. This set of simulations is based on Sgene2.

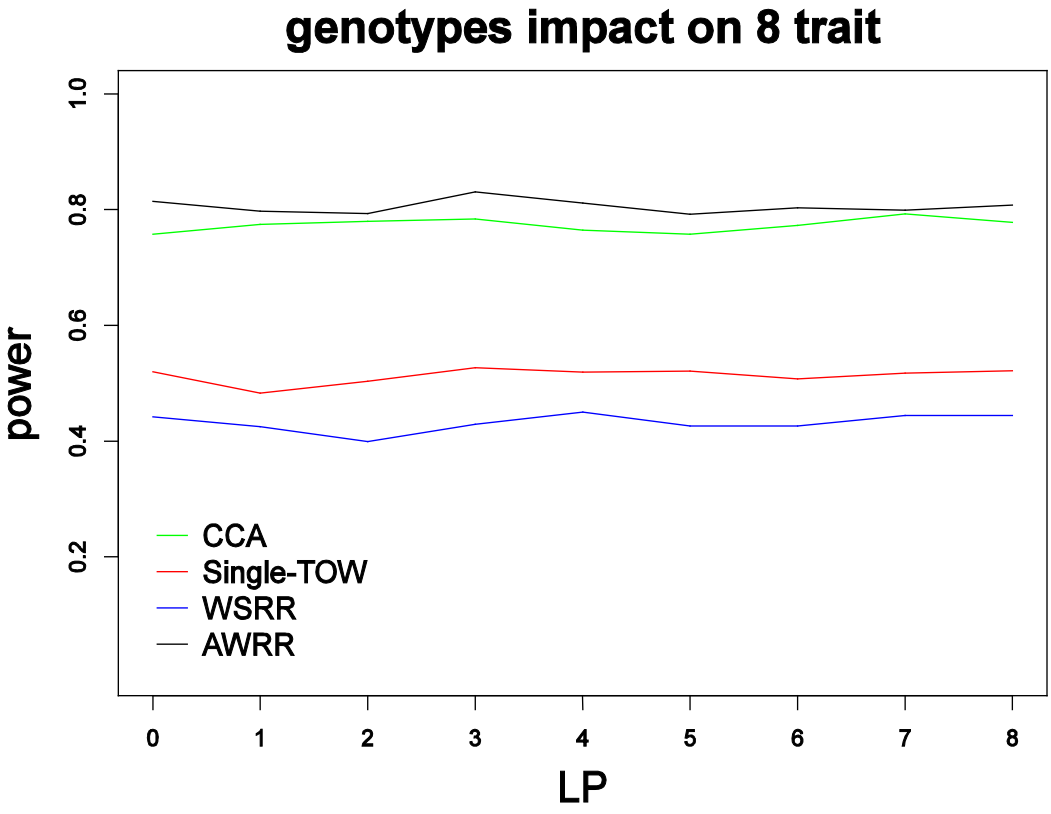


Figure B.2.1. Power comparisons of the five tests (SUM_SCORE, TATES, MHT, MHT-O and MANOVA) for the power as a function of the effect size (model 1). Sample size is 1000. Total number of traits is 20. The significance level is 5×10^{-8}. The number of replicates is 500. The number of permutations is 10^{8}.

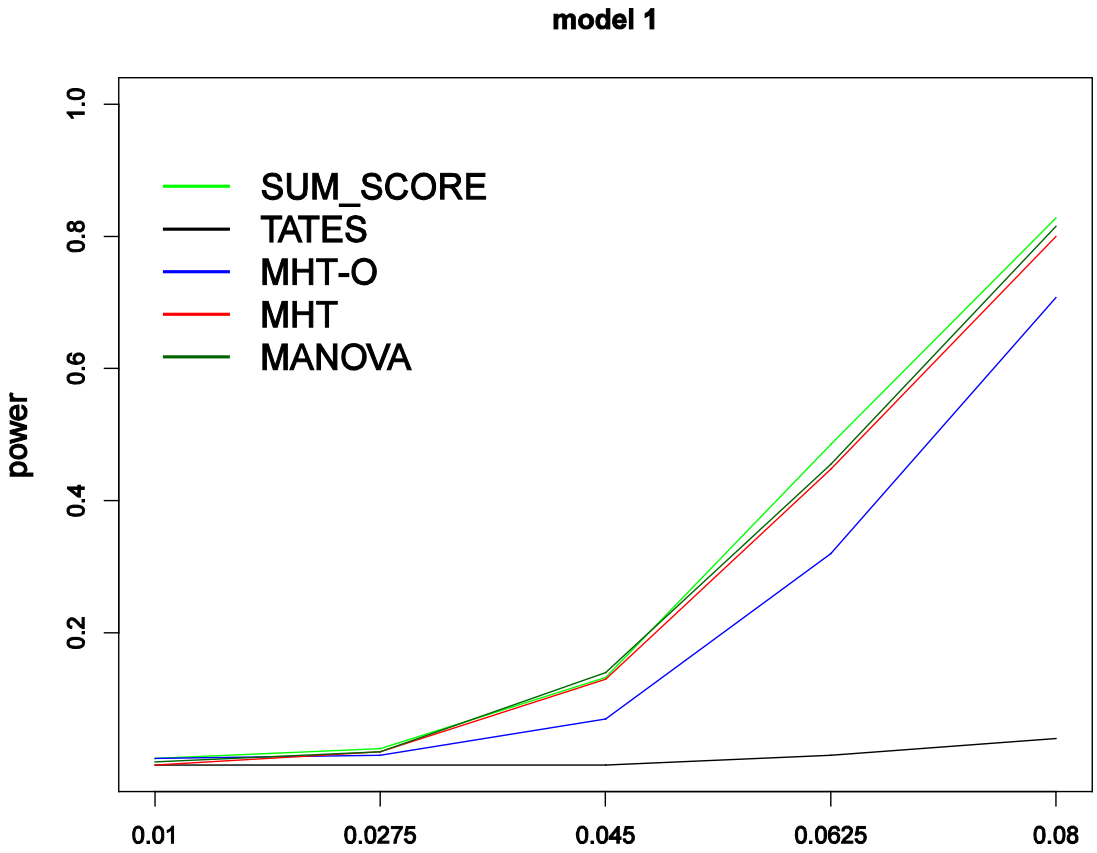


Figure B.3.1. Power comparisons of the six tests (Single-TOW, MSKAT, AWRR, MANOVA, GAMuT and TOWmuT) for the power as a function of within-factor correlation for models 1-5 and between-factor correlation for model 6 for 10 quantitative traits with covariates. The sample size is 1000. The percentage of the causal variants is 0.2. All causal variants are risk variants and $\rho = 0.5$ is for models 1-5. Heritabilities for models 1-6 are 0.05, 0.09, 0.08, 0.03, 0.03, and 0.06, respectively.

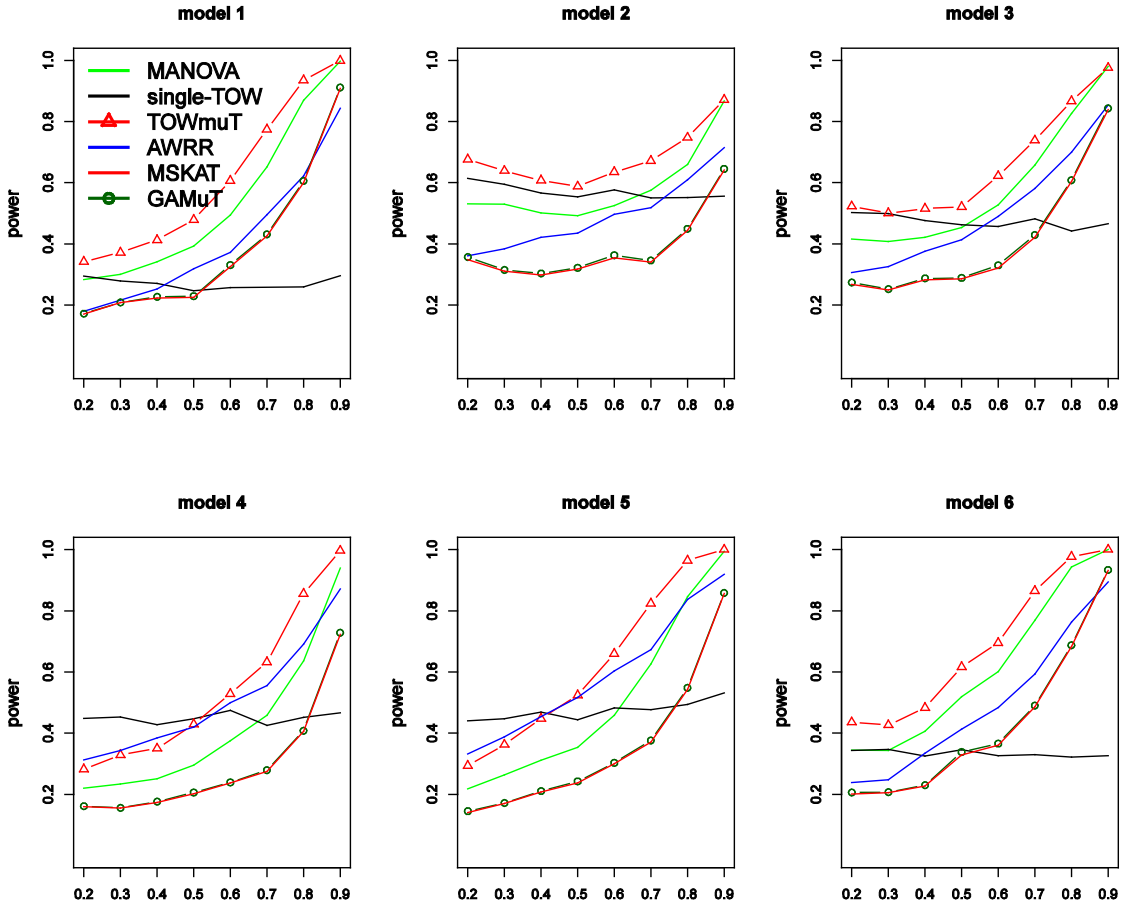


Figure B.3.1. Power comparisons of the six tests (Single-TOW, MSKAT, AWRR, MANOVA, GAMuT and TOWmuT) for the power as a function of the percentage of protective variants for 10 quantitative traits with covariates. The sample size is 1000. The percentage of the causal variants is 0.2. The between-factor correlation is 0.3 and the within-factor correlation is 0.7.